A FRAMEWORK FOR ESTIMATING PESTICIDE CONCENTRATIONS IN DRINKING WATER FOR AGGREGATE EXPOSURE ASSESSMENTS

An ILSI Risk Science Institute Working Group Report

ILSI Risk Science Institute
1126 Sixteenth Street, N. W.
Washington, D. C. 20036-4810 USA
Tel: (+1) 202 659-0074
Fax: (+1) 202 659-3859

ILSI Press
International Life Sciences Institute
1126 Sixteenth Street, N. W.
Washington, D. C. 20036-4810 USA
Tel: (+1) 202 659-0074
Fax: (+1) 202 659-3859

ISBN: 1-57881-073-6

Printed in The United States of America.

CONTENTS

ABOUT ILSI AND THE RISK SCIENCE INSTITUTE

The **International Life Sciences Institute (ILSI)** is a nonprofit, worldwide foundation established in 1978 to advance the understanding of scientific issues relating to nutrition, food safety, toxicology, risk assessment, and the environment. By bringing together scientists from academia, government, industry, and the public sector, ILSI seeks a balanced approach to solving problems of common concern for the well-being of the general public.

Headquartered in Washington, D.C., ILSI is affiliated with the World Health Organization as a nongovernmental organization and has specialized consultative status with the Food and Agriculture Organization of the United Nations.

ILSI accomplishes its work through its branches and institutes. ILSI's branches currently include Argentina, Australasia, Brazil, Europe, India, Japan, Korea, Mexico, North Africa and Gulf Region, North America, South Africa, South Andean, Southeast Asia, and Thailand, and a focal point in China. The ILSI Health and Environmental Sciences Institute focuses on global environmental issues. The ILSI Research Foundation includes:

ILSI Allergy and Immunology Institute

ILSI Human Nutrition Institute

ILSI Risk Science Institute

The ILSI Center for Health Promotion comprises the Physical Activity and Nutrition Program and the Micronutrient Deficiency Program/Project IDEA (Iron Deficiency Elimination Action).

The **ILSI Risk Science Institute (RSI)** was established in 1985 to advance and improve the scientific basis of risk assessment. RSI serves as a catalyst for consensus on complex scientific issues in risk assessment by facilitating discussion and cooperation among scientists from all sectors.

ACKNOWLEDGMENTS

The ILSI Risk Science Institute (RSI) has joined with the U.S. Environmental Protection Agency (EPA) in several cooperative agreements designed to answer questions in critical areas of risk assessment. This report resulted from an RSI initiative supported by cooperative agreements with the EPA Office of Pesticide Programs, Health Effects Division (including funding from the Environmental Fate and Effects Division), and the EPA Office of Water, Office of Science and Technology, Health and Ecological Criteria Division.

The Steering Committee for this activity included Dr. Timothy Barry (EPA Office of Policy Planning and Evaluation), Dr. Charles Crawford (U.S. Geological Survey), Dr. Elizabeth Doyle (EPA Office of Pesticide Programs), Dr. Carol Gotway-Crawford (Centers for Disease Control and Prevention), Dr. Raymond Layton (DuPont Agricultural Products), Dr. Keith Loague (Stanford University), Dr. L.D. McMullen (Des Moines Water Works), and Dr. Nicholas Poletika (Dow AgroSciences).

The 15 members of the expert Working Group that prepared this report were divided into three breakout groups to address monitoring, modeling, and data aggregation issues. The chairs of the breakout groups were Dr. Andrew Rogowski (Pennsylvania State University), Dr. Wendy Graham (University of Florida), and Dr. Douglas Crawford-Brown (University of North Carolina). Members of each breakout group are identified below and at the beginning of the chapters for which they had principal responsibility.

The RSI staff scientist responsible for this project was Mr. Thomas Brosnan. Mr. Brosnan convened and managed the activities of the Steering Committee and the Working Group and assembled and edited this report. At RSI, technical assistance and oversight were provided by Drs. Jeffery Foran and Stephen Olin, and administrative and logistical support for the project was provided by Eugenia Macarthy, Stephanie Carter, and Diane Dalisera.

Financial support from the U.S. Environmental Protection Agency and the ILSI Research Foundation is gratefully acknowledged.

BREAKOUT GROUP MEMBERS

BREAKOUT GROUP 1: THE ROLE OF MONITORING IN GENERATING ESTIMATES OF PESTICIDE CONCENTRATION DISTRIBUTIONS IN DRINKING WATER

Andrew Rogowski, Ph.D., Chair	Pennsylvania State University
Carol Gotway-Crawford, Ph.D.	Centers for Disease Control and Prevention
Russell Jones, Ph.D.	Rhone-Poulenc Ag Company
L.D. McMullen, Ph.D.	Des Moines Water Works
Thomas G. Sanders, Ph.D.	Colorado State University
James Wolf, Ph.D.	U.S. Environmental Protection Agency

BREAKOUT GROUP 2: THE ROLE OF MODELING IN GENERATING ESTIMATES OF PESTICIDE CONCENTRATION DISTRIBUTIONS IN DRINKING WATER

Wendy Graham, Ph.D., Chair	University of Florida
Charles Crawford, Ph.D.	U.S. Geological Survey
Conrad Heatwole, Ph.D.	Virginia Polytechnic Institute
Nicholas Poletika, Ph.D.	Dow AgroSciences

BREAKOUT GROUP 3: DATA AGGREGATION ISSUES

Douglas Crawford-Brown, Ph.D., Chair	University of North Carolina
Timothy Barry, Ph.D.	U.S. Environmental Protection Agency
Catherine Eiden	U.S. Environmental Protection Agency
Raymond Layton, Ph.D.	DuPont Agricultural Products
David Miller	U.S. Environmental Protection Agency

1. INTRODUCTION

With the enactment of the Food Quality Protection Act (FQPA 1996), the needs and uses of pesticide residue monitoring data for drinking water as required by the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) for pesticide registration have changed considerably. Under FQPA, the *unreasonable adverse effects on health or the environment* requirement of FIFRA and the requirement under the Federal Food, Drug, and Cosmetic Act that tolerances "protect the public health" have been brought together in a new general safety standard. FQPA requires that tolerances be established such that "there is a reasonable certainty that no harm will result from aggregate exposure to the pesticide chemical residue, including all anticipated dietary exposures and all other exposures for which there is reliable information."

Aggregate exposure assessments must consider multiple routes and sources of exposure for a single pesticide, including dietary exposure from food and water and residential exposure via dermal uptake and inhalation. Consideration of exposure through drinking water requires the use of "reliable information" on pesticide concentrations in potable ground and surface water.

In the assessment of pesticide concentrations in drinking water, a tiered process has been used to distinguish between those pesticides that do not pose an unacceptable risk, and therefore more detailed analysis is unwarranted, and pesticides that may require a full risk analysis via an aggregate exposure assessment. For those pesticides that fail the screening tiers, a preferred approach proposed for estimating the dietary portion of an aggregate exposure assessment is to combine a probabilistic drinking water exposure assessment with a probabilistic food exposure assessment, performed by a Monte Carlo analysis.

1.1. Charge to the Working Group

A number of issues need to be addressed in the long term to advance the science behind estimating pesticide concentrations in drinking water so that these estimates can be included in aggregate exposure assessments and cumulative risk assessments. The ILSI Risk Science Institute (RSI) undertook to address these issues through an expert Working Group in collaboration with the U.S. Environmental Protection Agency's (EPA) Office of Pesticide Programs and Office of Water.* RSI initiated this project by first convening a Steering Committee to help define some of these issues and provide guidance on the best ways to address them.

The Steering Committee suggested that guidance be developed for the collection or estimation of data (via modeling) that could be used to generate probability distributions of pesticide concentrations in drinking water for three scales of assessment: local, regional, and national. For each of these assessment scales, it is desirable to define data and modeling needs for relatively high, moderate, and relatively low degrees of confidence in the distributions. Distributions are needed for both new and existing pesticides and for drinking water derived from surface water and from ground water.

The following questions were proposed by the Steering Committee:

- What data are required for use in probabilistic aggregate exposure/cumulative risk analyses, and how can these data be collected?
- What role can modeling play in generating information/estimates that could be used to develop probability distributions of pesticide concentrations in drinking water for the three scales of assessment and the three confidence levels?
- How should monitoring and modeling data for drinking water be incorporated into an aggregate exposure analysis? Consider both the limitations of existing data and future, more complete data sets.

---

* This project is the second phase of an effort to evaluate the tools and methods available for estimating pesticide concentrations in drinking water. The first phase (ILSI 1998) focused on the identification of improvements to existing screening methods which could be implemented in the short-term.

A panel of 15 scientists was assembled and breakout groups of four to six scientists each were formed to deliberate each question. Each breakout group met twice in the fall of 1998 for 2 days. The chair of each group drafted a summary of his/her group's recommendations that was reviewed by the individuals in that group as well as by the scientists in the other two groups. Their findings constitute chapters 2, 3, and 4 of this report. Chapter 2 provides guidance on the design of a space-time sampling program to provide probabilistic data on pesticide residues in drinking water at the national, regional, and local scales. Chapter 3 discusses the role of modeling in generating estimates of pesticide concentration distributions in drinking water. Chapter 4 explores various concepts to consider when incorporating drinking water data into an aggregate exposure assessment.

1.2. Scope and Definitions

1.2.1. Use Regions, Assessment Areas, and Scales

A *use region* is the combined area that receives application of a particular pesticide for a specific crop or noncrop use or for a set of such uses. A use region theoretically can consist of areas in several noncontiguous counties, states, or watersheds (Figures 1.1 and 1.2). An *assessment area* is defined as any geographic region that one might choose to do a risk assessment on. An assessment area could be defined, for example, on the basis of hydrological boundaries; common climate, soil, farming, and other ecological characteristics; or other designations (Figures 1.3–1.6).

When such definitions are used, *for any given risk assessment* a common definition needs to be applied consistently, and care must be taken in combining regional distributions of pesticide concentrations in drinking water to form a national distribution. Presumably, the collection of all regional distributions would form the national distribution, but this process could lead to errors in the national distribution if dissimilar concepts of the term "region" are combined. For example, regions defined by farming practices and regions defined by hydrological boundaries may contain common water supplies. Combining these two types of regions to form a national distribution would lead to "double counting" of those water supplies common to the two categories of regions.

Depending on the scope and purpose of the risk assessment, it is conceivable that complete probability distributions of pesticide concentrations in drinking water might be needed for one or more of the following assessment scales:

- local (temporal data needed),
- regional (spatial and temporal), and
- national (spatial and temporal).

A local scale implies some type of site-specific as-

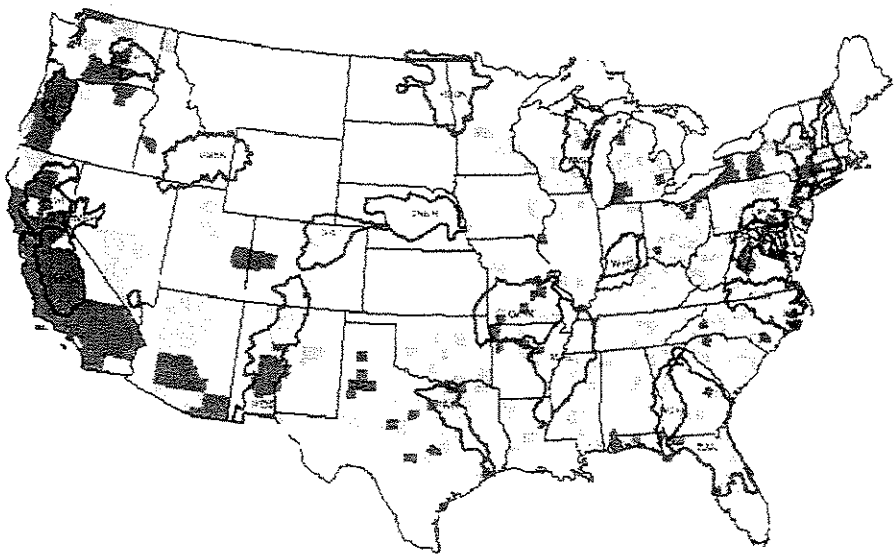

Figure 1.1. Example of a use region for pesticide *X* in 1998. Modified from USGS http://water.wr.usgs.gov/pnsp/rep/carbo/map13.html.

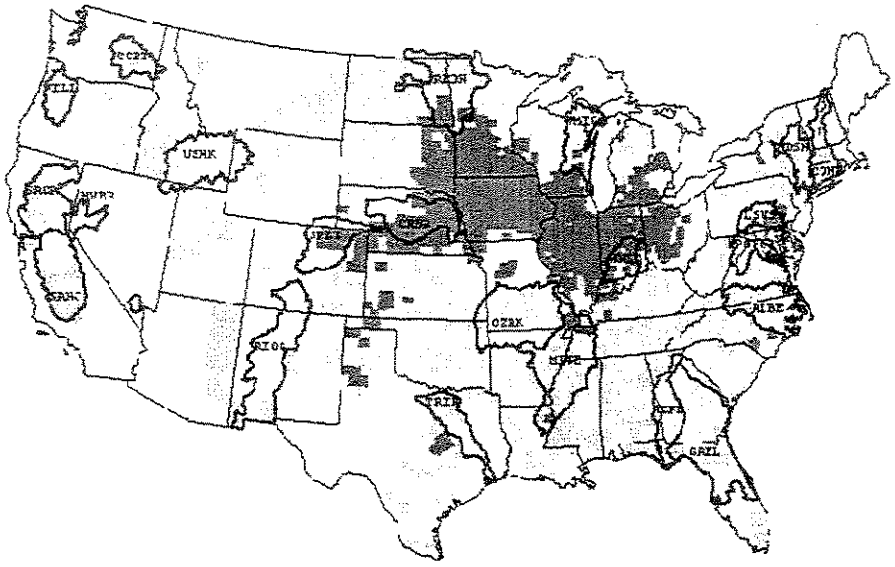

Figure 1.2. Example of a use region for pesticide *Y* in 1998. Modified from USGS http://water.wr.usgs.gov/pnsp/rep/carbo/map6.html.

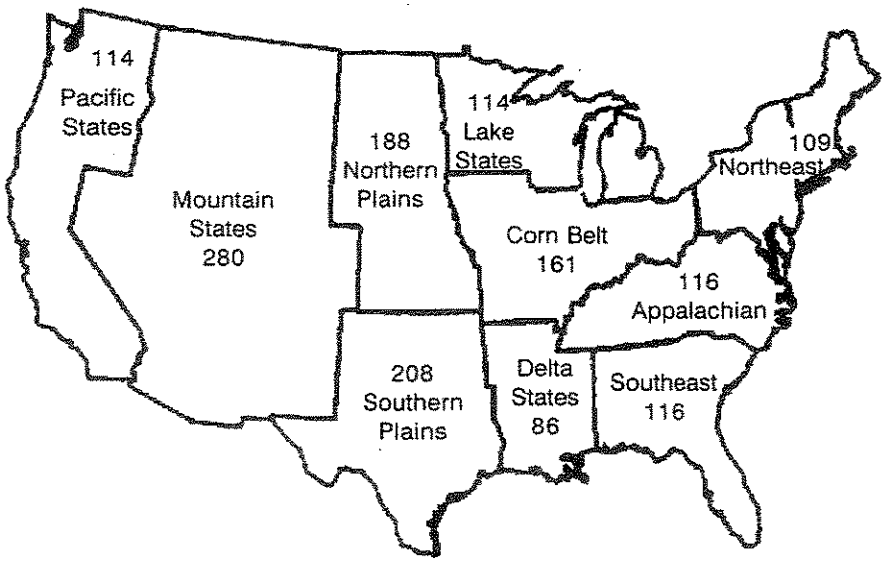

Figure 1.3. Farming regions: nonfederal land (millions of acres) in each farming region (USDA 1989).

sessment of an individual supply. A regional-scale assessment might include a number of water supplies within the use region of an assessment area, whereas the national-scale assessment incorporates some aggregation of water supplies from several assessment areas.

Although the objective of this report is to provide guidance to ensure that future drinking water data are collected so that the full statistical (probability) distribution can be characterized, it is recognized that it often may be desirable to characterize some percentile at the upper tail of the distribution, either exclusively or particularly well; i.e., sampling would be targeted at a selection of supplies that are for various reasons most likely to contain the highest pesticide concentrations. Note that of the scales described above, local sampling would most likely be targeted at a most vulnerable supply, if such a supply could reliably be identified a priori. However, defining the one most vulnerable supply in an assessment area based on site characteristics and pesticide loading and fate and transport properties is conceptually and practically very problematic, and the designation of "most vulnerable" could change as weather, pesticide use, and other dominant factors change over the years in the assessment area. Therefore, although it is conceivable that a full pesticide

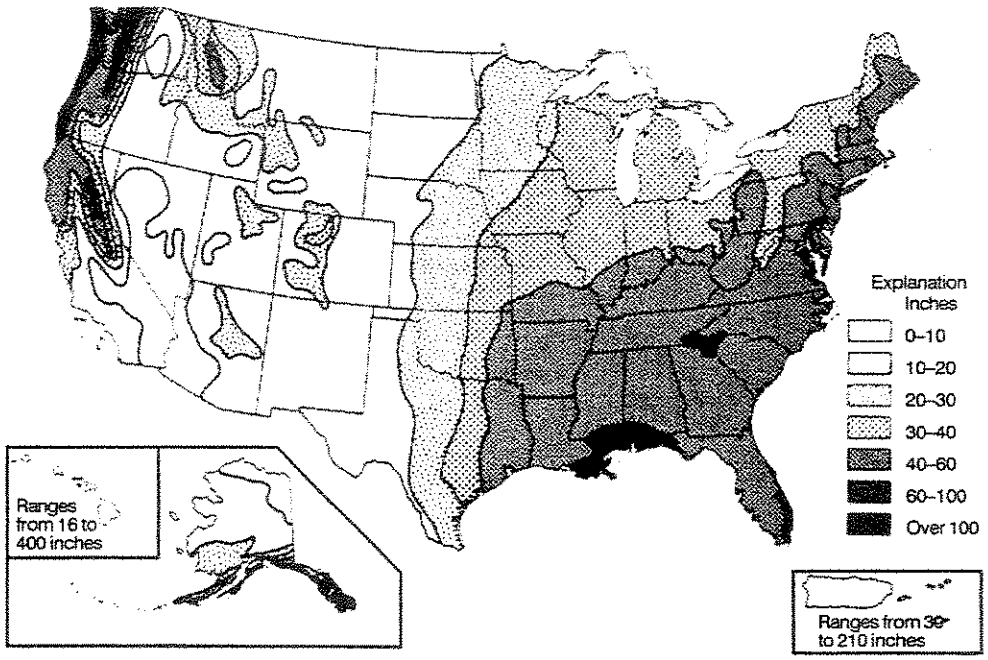

Figure 1.4. Average precipitation regions in the United States (USDA 1989)

Figure 1.5. Water resource regions in the United States (USDA 1989)

concentration distribution from a single water supply might need to be characterized, it is the least likely scenario to be assessed. More often, for risk assessments that need to define the upper tail exclusively or particularly accurately, sampling of a *class of vulnerable supplies* is a more defensible choice. In a modeling exercise, however, a single site can be parameterized based on prior knowledge to simulate a known level of vulnerability based on site characteristics and climate, and simulation of a class of sites may be unnecessary. This topic is discussed further below.

1.2.2. Types of Drinking Water Data

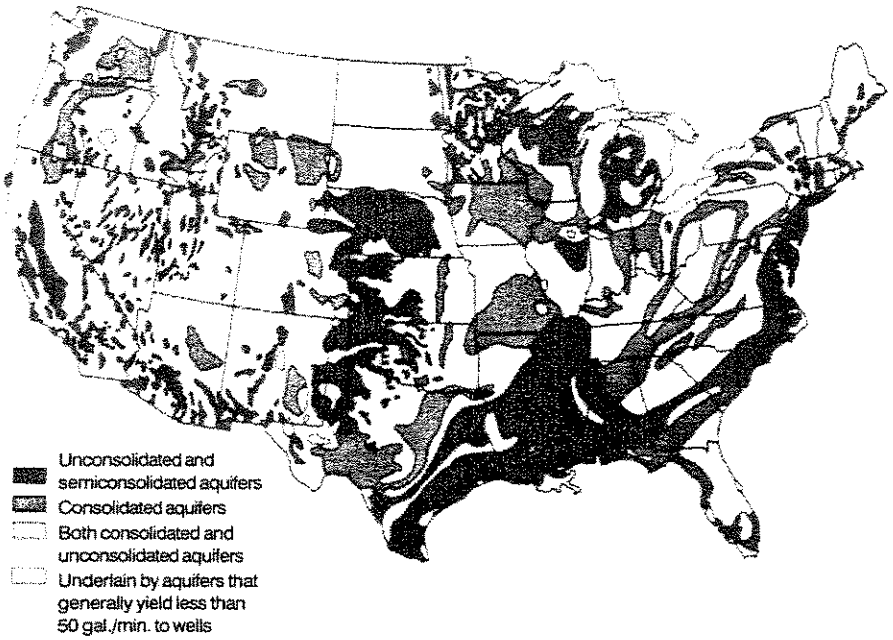

Figure 1.6. Major aquifer regions of the United States. More than 80% of all water pumped from wells comes from gravel and sand aquifers. If the aquifer is close to a water course, infiltration from the river normally replaces the water withdrawn from wells (USDA 1989).

Needed

To perform aggregate exposure assessments for a pesticide, information on pesticide exposure from drinking water must be expressed in terms that can be combined with other dietary exposure estimates. Dietary risk assessments are developed for three types of effects: acute noncancer effects (single dose), chronic noncancer effects, and cancer. Initial estimates of acute dietary exposures from food typically use highest average field trial data for single-serving commodities or 95th percentile monitoring data for blended commodities. In subsequent (higher) acute dietary assessment tiers, the entire range of pesticide concentrations obtained from field trials may be used to generate probabilistic acute dietary assessments via Monte Carlo analyses. To estimate chronic exposures for the food portion of the dietary exposure, average field trial data or, when available, average monitoring data are used.

To perform a scientifically defensible probabilistic aggregate exposure assessment that incorporates acute and chronic drinking water exposure data from an individual supply, reliable information on the full temporal distribution of pesticide concentration is needed. This temporal distribution can be used to estimate peak concentrations (via short averaging periods, such as 24 hours, or as a maximum instantaneous value or a maximum annual daily value) and long-term average concentrations (e.g., annual or lifetime exposures; e.g., see Figure 2.1 in Chapter 2). To estimate regional or national acute and chronic exposure, distributions over both space and time are necessary.

1.2.3. Levels of Relative Accuracy

For each of the three assessment scales, sampling plans are needed to construct a probability distribution at each of the three levels of relative accuracy: high degree, moderate degree, and low degree. The degree refers to an expected level of accuracy that a generated pesticide concentration distribution would have, based on the quality and quantity of data provided by a particular monitoring or modeling strategy, prior information, and statistical assumptions. It is understood that increased levels of accuracy require more and higher-quality sampling data and modeling analyses. Conversely, decreased levels of relative accuracy may be achieved with less stringent data requirements, but need to be supplemented by statistical assumptions and prior information.

1.2.4. Sampling of Tap Water Versus Intake Water

Ideally, samples for measurement of pesticide residues in drinking water should be collected at the tap, since this is the water that is actually being consumed. If an adequate representative and unbiased data set of tap water concentrations is available for the assessment area, those data are appropriate for use in distributional analyses for aggregate exposure assessments. However, tap water is a time-varying composite of source water(s) at the intake that is further subjected to potentially varying treatment effects, distribution allocation, and mixing in transit. Designing a generic tap water sampling scheme that incor-

porates these factors was considered beyond the scope of this project.

Although the majority of regulatory data sets are currently being collected from tap water, the authors of this report noted that there are advantages to collecting data from intake water. These intake samples serve as a conservative estimate of pesticide residues in drinking water. The estimates can be made more realistic by considering the effect of treatment. This could be done either by modeling the effectiveness of treatment in reducing residues or by comparing intake and tap analyses. Ideally, if resources are available, both intake and tap samples could be collected, and tap samples analyzed when the intake had pesticide residues above the limit of quantification or a designated level of concern.

1.2.5. Sources, Types, and Population of Drinking Water Supplies

Under the Safe Drinking Water Act (SDWA), EPA is responsible for regulating more than 200,000 distinct public water systems, each serving at least 25 people. There are 18 major water resource regions located in the coterminous United States. More than 120 million people derive their drinking water from surface water supplies such as rivers, streams, impoundments, catchments, and springs (Solley et al. 1993). Of this population, more than 98% derive their drinking water from public supplies and 1,660,000 derive drinking water from domestic supplies. Municipal surface water systems that use large rivers for their drinking water supply represent the greatest segment of this population.

Approximately 126,750,000 people derive their drinking water from ground water supplies. Of these, 86,090,000 derive their drinking water from public supplies, whereas 40,660,000 people use domestic supplies (Solley et al. 1993). Approximately 95,000 drinking water wells serve about 40,000 community water systems and about 13 million rural domestic wells in the United States (EPA 1990). Overall, approximately 12% of the U.S. population (28 million people), in mostly rural areas, obtain their drinking water from unregulated individual domestic wells, springs, small ponds, and catchments (U.S. Department of Agriculture 1989, EPA 1990).

References

International Life Sciences Institute. 1998. Assessment of methods to estimate pesticide concentrations in drinking water sources. Washington, DC: ILSI Risk Science Institute

Solley WB, Pierce RR, Perlman HA. 1993. Estimated use of water in the United States in 1990. US Geological Survey Circular 1081. Washington, DC: US Geological Survey

US Department of Agriculture. 1989. The second RCA appraisal: soil, water, and related resources on nonfederal land in the United States: analysis of conditions and trends. Washington, DC: US Government Printing Office

US Environmental Protection Agency. 1990. National Pesticide Survey: summary results of EPA's National Survey of Pesticides in Drinking Water Wells. EPA 570990NPS5. Washington, DC: USEPA Office of Water, Office of Pesticides and Toxic Substances

2. THE ROLE OF MONITORING IN GENERATING ESTIMATES OF PESTICIDE CONCENTRATION DISTRIBUTIONS IN DRINKING WATER

Breakout Group 1: Andrew Rogowski, Pennsylvania State University (chair); Carol Gotway-Crawford, Centers for Disease Control and Prevention; Russell Jones, Rhone-Poulenc Ag Company; L.D. McMullen, Des Moines Water Works; Thomas G. Sanders, Colorado State University; and James Wolf, U.S. Environmental Protection Agency

2.1. Introduction

Breakout group 1 was charged with developing a monitoring strategy to answer this question: *What types of drinking water data should be collected, and how should they be collected, to ensure that they are compatible with other exposure data collected and used in an aggregate exposure analysis?*

The following questions were provided as a framework for the discussion:

- What types of data are needed to generate a defensible probability distribution? Consider, for example, acute/ chronic and temporal/ spatial components.
- What are the size and nature of the data set needed, including the minimum data set required to have high, moderate, and low degrees of confidence in probability distributions generated for each of the three assessment scales?
- What are the roles of prior screening levels and of directed and statistically based monitoring?
- What is the relationship between temporal and spatial variability versus the size of the assessment area?
- What ancillary data need to be collected? For example, when, where, and how much pesticide has been applied? When, where, and how much has it rained? What are the hydrogeologic setting, climate, land use, etc.? How might these data be helpful, and where are they available?

These questions were addressed by providing guidance on the development of a space-time sampling design to monitor pesticide concentrations at the national, regional, and local scales. This is a major undertaking that requires additional extensive information, computer resources, discussion, and research. This report gives some preliminary ideas in support of this design task, identifies key terms, and ultimately provides some initial guidelines regarding sampling designs for monitoring pesticide concentrations in drinking water that can form a basis for more intensive work and discussion. The authors stress that these recommendations should be viewed as initial guidance; actual monitoring design requirements will change based on the purpose and scope of the risk assessment, site and pesticide properties, and other factors. Surface water supplies and ground water supplies are discussed separately.

Best professional judgment. The breakout group developed this initial guidance using their collective "best professional judgment." The underlying rationale for the design was based on a statistical probability sampling with appropriate stratification, but this rationale was modified at times to reflect the real cost or practical and logistical constraints. The process may be understood as a systems approach to selection among alternatives (Saaty and Alexander 1981). In general, any systems approach is hierarchic and proceeds from the particular to the general. The quantified judgments on pairs of activities are usually represented by a matrix of relative weights arrived at by comparing the relative importance of adjacent pairs, where experience and collective judgment favor the specified approach. The decision-making process created the first row of the matrix (i.e., high accuracy). When judgments appeared to be consistent, requirements for the remaining rows of the matrix were determined by relaxing the requirements and substituting statistical assumptions and prior knowledge for data. Thus, the assessment described here starts with projecting sampling requirements for a local site over time at a relatively high level of accuracy and proceeds to requirements in time and space at a regional and national scale, with a relatively low level of accuracy.

The recommendations were tempered by the high cost of analysis, the desire to avoid collecting large numbers of nondetect samples, practicality, and applicability of the proposed sampling strategy. They strike a balance between the number of samples optimally required to describe pesticide distributions at a sufficiently high level of accuracy and the practical feasibility of implementing the advocated sampling plan. In that sense, the recommendations, especially at the high-accuracy level, represent a minimum number of samples required to develop a statistically sound distribution. To provide some substantiation for these choices, the recommendations were checked with real data sets. The results are discussed in the Appendix to Chapter 2 and in Section 2.2.3 below.

However, it must be recognized that a number of issues require further consideration. First, what should be the goal for sample size determination? Should it be to capture a true upper-tail percentile? To span a given percentage of an underlying distribution with known confidence bounds? To estimate an entire distribution with specified confidence bounds? Or to estimate a particular percentile with a specified degree of confidence? These different goals potentially can result in very different requirements for sample size. Second, what is, or should be, the role of randomized, stratified, probability-based sampling? Statisticians would agree that this is crucial, but implementing such a strategy may be impractical and the resulting analyses can be complicated.

2.2. Sampling Plans for Surface Water Supplies

Surface water monitoring studies conducted to date are often not adequate to meet the "reliable information" clause necessary to perform drinking water exposure assessments as required under the Food Quality Protection Act (FQPA). Notable limitations were that sampling points did not represent drinking water sources and that sample designs were generally not spatially and temporally adequate for a probabilistic drinking water exposure assessment. Other limitations included inadequate focus on major pesticide use regions, over- or underrepresentation of vulnerable locations, and sampling that did not consider relative timing of application and runoff events.

Based on best professional judgment, Tables 2.1 and 2.2 provide guidance for each level of accuracy on the minimum number of samples to collect over time that should be considered to describe a distribution at an individual supply; for the regional and national scales, the minimum number of water supplies required to develop a spatial distribution in time for each assessment area is provided. Sampling plans at a local, regional, and national scale were conceived as being separate and independent of one another. Guidance on how to choose which water supply sites to sample is provided in Section 2.2.4. Note that the surface water guidance can also be applied to ground water supplies under the influence of surface water. These plans are proposed as general guidance to developing sampling plans. For the design of specific plans, modifications may be appropriate based on a variety of factors such as pesticide properties, agricultural practices, environmental conditions, and characteristics of specific water supplies.

2.2.1. Sampling Plan to Estimate Peak Concentrations (as Daily Means)

Table 2.1 is a suggested sampling matrix for estimating the distributions of daily mean peak concentrations of a pesticide in surface drinking water supplies. The concept is to sample systematically over time, once per day, so that a distribution of daily values, both high and low, can be developed from which a "daily mean peak" can be estimated.

The relative merits of "grab" sampling versus continuous or even composite sampling in time were considered. Rather than attempting to obtain peak concentration values by sampling continuously, the exposure statistics would be better served if a distribution of values was obtained with some probability that it contained a peak concentration. Grab samples taken systematically with a random start on a seasonal basis will in time describe the shape of the entire distribution, including the peak values.

2.2.1.1. Sampling Plan for a Local-Scale High-Accuracy Assessment

As noted earlier, monitoring to characterize a distribution at an individual supply would most likely be performed on a supply that was believed to represent the most vulnerable supply in the assessment area. Actually defining any one supply as the most vulnerable in an assessment area is highly problematic, so it is reasonable to assume that there may be other reasons why a risk assessor would want to characterize pesticide distributions at an individual supply. Therefore, sampling requirements were defined for deriving a distribution of concentrations from an individual supply.

Sampling every day to characterize a distribution of concentrations over time at a single supply would represent a maximum-level-of-effort sampling strategy. However, this is very costly and unnecessary if an alternative statistical probability sampling approach can be used effectively (Olsen et al. 1999). Thus, options to capture the temporal distribution centered on using appropriate statistical sampling procedures to obtain a representative distribution of concentrations in time. Moreover, given the known seasonality of pesticide occurrence in waterways (Larson et al. 1996), a more efficient monitoring strategy involves the use of a seasonally stratified sampling plan.

Table 2.1. Sampling matrix for daily mean (peak) pesticide concentration in surface water supplies at three spatial assessment scales and three levels of accuracy

	Spatial assessment scale		
Prediction accuracy level	**National assessment (multiple regions with multiple water supplies)**	**Regional assessment area (multiple water supplies)**	**Local assessment (single water supply)**
High	10–15 water supplies per region ~ 33 samples per supply per year Supplies aggregated over regions	20–30 water supplies per assessment area ~33 samples per supply per year	33 samples per water supply per season for 3 seasons ≈ 100 samples per year
Medium	5–10 water supplies per region ~33 samples per supply per year Supplies aggregated over regions + assumptions about the distribution	10 water supplies per assessment area ~33 samples per supply per year + assumptions about the distribution	10 samples per water supply per season for 3 seasons ≈ 30 samples per year + assumptions about the distribution
Low	Existing data[a] for regions aggregated nationally + many assumptions about the distribution	Existing data for the same assessment area + many assumptions about the distribution	Existing data from nearby water supplies + many assumptions about the distribution

[a] Public water supplies are required to take four samples and four replicates per year.

The seasonality of pesticide occurrence in waterways does not necessarily have a calendar connotation of spring, summer, fall, and winter. Rather, seasonality varies among supplies, based on geographic region, local climate, relative timing of rainfall events and applications, characteristics and intensity of use, and pesticide fate and transport properties. Surface supplies in many areas have three pesticide occurrence seasons. The first is the dormant season in late winter/early spring before pesticide application, when concentrations in surface supplies are very low to nondetectable. The second is spring/summer, when concentrations may be at their peak. The third is the late summer/fall season, when concentrations are relatively low again. (Note that this discussion applies more directly to "rain-fed" agriculture than to western irrigated agriculture.)

In attempts to develop a statistical distribution, a sampling cycle needs to coincide with a pesticide concentration history that represents a recurring time frame. For surface water, the sampling cycle, if representative of a normal precipitation pattern, generally covers 12 months. Typically, pesticide concentrations in surface water and

Table 2.2. Sampling matrix for annual time-weighted average pesticide concentration in surface water supplies at three spatial assessment scales and three levels of accuracy

	Spatial assessment scale		
Prediction accuracy level	**National assessment (multiple regions with multiple water supplies)**	**Regional assessment area (multiple water supplies)**	**Local assessment (single water supply)**
High	10–15 water supplies per region 12–24 samples per supply per year Supplies aggregated over regions	20–30 water supplies per assessment area 12–24 samples per supply per year	12–24 samples per year collected monthly to twice monthly, with longer sampling interval in nondetect seasons
Medium	5 water supplies per region 6–12 samples per supply per year Supplies aggregated over regions + assumptions about the distribution	10 water supplies per assessment area 6–12 samples per supply per year + assumptions about the distribution	6–12 samples per year + assumptions about the distribution
Low	Existing data[a] for regions aggregated nationally + many assumptions about the distribution	Existing data for the same assessment area + many assumptions about the distribution	Existing data from nearby water supplies + many assumptions about the distribution

[a] Public water supplies are required to take four samples and four replicates per year.

in ground water under the influence of surface water are highest if a precipitation event follows pesticide application and are lowest during the dormant season (Barbash and Resek 1996, Larson et al. 1997). To enhance interpretation of the occurrence of concentration peaks and to verify anomalous situations, it is recommended that precipitation data be acquired in the assessment area at the time of sampling and compared with the historical norm for that area.

To fully characterize an individual surface water supply (or ground water under the influence of surface water) in time, a minimum of 33 samples should be collected per season (based on best professional judgment). For the three pesticide occurrence seasons typical of many areas, this would result in approximately 100 samples collected per year. But how should these 33 samples per season be spaced in time to ensure that the peak concentration values are captured?

In a statistical sampling approach, characterizing a distribution means adequately characterizing the entire distribution, including both the upper and the lower tails, as well as the middle of the distribution. If there was no prior knowledge of a distribution, the 33 samples could be allocated evenly over each of the three seasons, with a random start and fixed sampling times. From a practical point of view, however, if the risk assessor's interest is primarily in the upper tail, a distribution can be represented by fewer samples in the lower tail and in the middle and more in the upper tail, as long as they are appropriately weighted. Because most literature suggests that pesticides are more commonly detected in surface water during the application season and are often nondetectable in the dormant season (Larson et al. 1997), a sampling plan that requires 33 samples allocated evenly over each season may provide many nondetects in the dormant season, and could be viewed as wasteful of resources. The suggested alternative strategy would be to ensure characterization of the whole distribution, with additional emphasis on the upper tail by weighting the sampling based on seasonal stratification. Thus, more samples (i.e., more than 33) would be collected during the season of primary application, with less than 33 samples collected in other seasons. Under this scenario, all seasons would have a known finite chance of being sampled, but the dormant season would be sampled least. Oversampling during high-risk seasons is statistically acceptable as long as there also is an adequate probability of selection of samples from water supplies in every other season.

2.2.1.2. Sampling Plan for a Regional-Scale High-Accuracy Assessment

A regional assessment area was envisioned as a more or less homogeneous collection of individual water supplies within a watershed or across several watersheds within a pesticide use region. Note that a regional-scale assessment requires distributions over both space and time. Distributions over time for a subset of individual supplies are first generated, and then distributions of peaks and long-term means of pesticide concentrations for water supplies in an assessment area within a use region can be estimated (Figure 2.1).

To develop high-accuracy distributions for a regional-scale assessment, it is recommended that sampling 20–30 water supplies per assessment area should be considered, with up to 33 samples collected per year from each, for a total of approximately 900 samples per assessment area per year (Table 2.1). For assessments designed to characterize only the upper end of the probability distribution, fewer supplies might be needed. Timing of a sampling campaign and other details of the collection of these 33 samples at each of the water supplies should follow the plan outlined above for the local supply. The question of how and where to choose the 20–30 water supplies to sample in an assessment area is dealt with in more detail in Section 2.2.4 (Sampling Locations) below.

2.2.1.3. Sampling Plan for a National-Scale High-Accuracy Assessment

The national-scale assessment is perceived as an aggregation of water supplies sampled from several regional assessment areas. To accomplish this with a high level of accuracy, it is suggested that 10–15 water supplies per assessment area be sampled, with 33 samples collected per year at each site (Table 2.1). Just as the number of samples per water supply is decreased at a regional scale when more than one water supply is sampled, the number of water supplies that needs to be sampled at a national scale decreases when more than one assessment area is involved.

2.2.1.4. Sampling Plan for a Local-, Regional-, or National-Scale Medium-Accuracy Assessment

When water supplies are monitored for pesticide residues, there often may not be enough data to construct a probability distribution at a high level of accuracy, or there may not be enough resources to collect these data. A medium-accuracy assessment can be achieved with fewer samples by relying on statistical assumptions about the distribution and on prior information such as existing data. Statistical assumptions include assumptions about the shape of the distribution, for example, log-normality, and uncertainty around the predictions. Existing data include supplemental historical data sets that describe the spatiotemporal concentration of pesticides in the environment relating to the same area where the monitored water

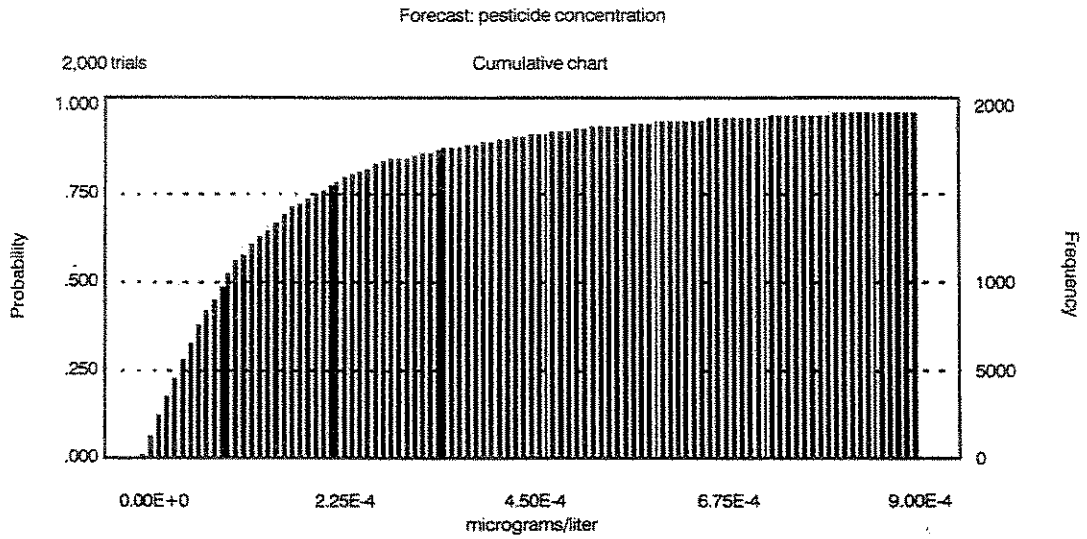

Figure 2.1. A hypothetical cumulative distribution function for the concentration of a pesticide in one or more water supplies. It might, for example, be the distribution of concentrations over time at an individual water supply, or of peak or median concentrations across several water supplies within a use region. The x axis displays the concentrations. The y axis displays the fraction of time (for an individual supply) or fraction of water supplies (for regional or national assessments) with a peak less than or equal to the values on the x axis. The fraction of time or water supplies exceeding a given peak concentration can be found by subtracting the y axis values from 1.0.

supplies are located. Such data may have been acquired as part of routine monitoring or exploratory studies by state or federal agencies, commercial companies, or academic research scientists on a local, regional, or national scale (e.g., Gilliom et al.1995, Larson et al.1997, U.S. Environmental Protection Agency [EPA] 1990). Related data sources include compiled measurements of treated tap water that each public water supply is required to make four times per year (EPA 1990).

Before a regional or national medium-accuracy sampling assessment can be performed, these supplementary data sources in particular need to be examined to broadly identify the spatial extent of potential contamination. To design a viable monitoring strategy, certain assumptions associated with such preexisting historical data sets may also need to be made, for example, how to extrapolate lower and upper tails of the actual distribution or how to interpolate between measured values. Although information derived from such existing distributions and assumptions may have different levels of uncertainty associated with them, they can be used as a guide in designing a sampling plan. Guidance on the extrapolation and interpolation of data is contained in Chapter 4.

2.2.1.5. Sampling Plan for a Local-, Regional-, or National-Scale Low-Accuracy Assessment

Using existing data and many assumptions about the shape of the distribution, one can generate probability distributions with a low level of relative accuracy. Existing data could be from a location similar to currently sampled drinking water supplies, for example, data from the U.S. Geological Survey's National Water Quality Assessment program (USGS NAWQA [Gilliom et al. 1995]). These data need not be for the supply itself and may have been collected for another reason, in another location, and at another time. The low-accuracy estimates could use, for example, as the first approximation, the distributions based on measurements of pesticide concentration in drinking water sampled four times per year, which are already available from different water treatment plants throughout the country. However, because it would be desirable to have the supporting databases as uniform and unbiased as possible even at this low accuracy level, stricter and more uniform data collection guidelines should be enforced nationally.

2.2.2. Sampling Plan to Estimate Long-Term Mean Concentrations (as Annual Time-weighted Averages)

Adherence to the sampling strategy proposed for daily (peak) concentrations in Table 2.1 should also provide adequate data for a reasonable representation of long-term average concentration of pesticides in drinking water. Where estimates of daily means are not required, a reduced level of effort as detailed in Table 2.2 will characterize the central tendency of a distribution, which is suitable for long-term average (chronic) exposure estimates. Test results in the Appendix to this chapter provide further insights into the appropriate numbers of samples to

Table 2.3. Total number of samples recommended for the estimation of the daily mean (peak) and annual time-weighted average concentration of pesticides in surface water supplies at three spatial assessment scales and two levels of accuracy

	Daily			Annual[a]		
Confidence	National	Regional	Local	National	Regional	Local
High	~2500	1000	100	~1800	720	24
Medium	~1500	100	30	~300	120	12

[a]These numbers could be doubled if weekly sampling is adopted.

estimate various percentiles of a distribution (see also Section 2.2.3). As described previously, however, if the goal is to avoid the collection of excessive numbers of samples with nondetectable residues, sampling can be weighted such that more samples are collected during the runoff season and fewer collected in the dormant season.

The total numbers of samples suggested for the daily and annual assessments are illustrated in Table 2.3. National estimates assume a composite of up to five assessment areas. Water supplies sampled for long-term averages may also be drawn from those sampled for a daily peak assessment. For example, 12–24 samples collected monthly to twice monthly (Table 2.2) could be drawn from the samples already taken from selected local supplies (Table 2.1). Similarly, 20–30 water supplies per assessment area sampled for long-term averages at the rate of 12–24 samples per year could include subsampling of supplies previously used to estimate peak daily concentrations, and the same again at the national scale.

If weekly sampling was substituted for the twice monthly (Table 2.3), the recommended number of samples to estimate the long-term average (chronic) exposure would approximately double.

2.2.3. Statistical Evaluation

The sampling strategies for surface water discussed above and summarized in Tables 2.1–2.3 were derived from the best professional judgment of the breakout group tempered by considerations of practicality and cost. To provide some perspective on these strategies, a statistical analysis was performed to evaluate these recommendations in terms of confidence in estimating specific percentiles of the pesticide concentration distributions (Appendix). The analyses of surface water data consider only the scenarios for the local high-accuracy assessment in the daily sampling matrix. The scope of analysis is presently limited to the question of the adequacy of the number of samples to collect at a site. No analysis is made (for surface water supplies) of the adequacy of the number of supplies to sample in the regional and national surveys. Initially the analyses were conducted on data from a single water supply to compare the proposal to use 33 samples per season (Table 2.1) with the actual samples needed to estimate the upper percentiles of the distribution. The results indicated that, subject to assumptions detailed below, the sample size recommendations were minimally adequate, so that the estimation of the extreme upper percentiles would require a much larger number of samples or a different sampling strategy.

As discussed in the Appendix to this chapter, there are many different ways to impose statistical criteria for sample size determination. One is to use a nonparametric estimator for a given percentile and then require that this estimator be "close" to the true value with some specified degree of confidence. The use of this nonparametric estimator gives some generic guidance on the minimum number of samples needed to estimate specific percentiles, since the nonparametric procedure is ill defined and will not work if the number of samples is too small. Thus, with reference to the discussion in the Appendix, at least 20 observations are needed to estimate the 95th percentile of a distribution, whereas at least 100 samples are needed to estimate the 99th percentile.

The use of *tolerance intervals* is another way to gauge how representative the sample values are. With this approach, it is possible to predetermine a sample size that will, with a specified probability, cover a desired percentage of the distribution of population values. This coverage is the tolerance interval. The first example in the Appendix targets a sampling strategy to capture the upper end of a distribution. Table A1 in the Appendix gives the smallest sample size needed to ensure that the largest value in the sample exceeds the pth percentile of the population values. This is a necessary requirement if one is to reliably estimate the pth percentile of a population distribution. For example, 45 samples are needed to have a 90% probability that the maximum value in a sample exceeds the 95th percentile of the population distribution. Table A1 demonstrates one way to provide some perspective on the adequacy of the best professional judgment used to construct Tables 2.1–2.3 above. For example, the 33 samples per season suggested for the local supply at the high-accuracy level appear adequate to ensure that the maximum in the sample exceeds the 90th percentile with almost 98% probability, exceeds the 95th percentile with between 75% and 90% probability, or exceeds the 98th percentile with almost 50% probability.

Note that it is necessary for the sample maximum to

exceed the specified percentile of interest for estimating this percentile, but no statements can be made about the accuracy of the resulting estimates. Such statements would require consideration of some type of confidence interval for the percentile estimates. Such intervals are given for the nonparametric approach to percentile estimation in the Appendix. However, using these intervals to analytically provide a formula for required sample sizes (as is often done in the case of estimating a mean from a Gaussian distribution) is very difficult. Thus, the sampling strategy recommendations were evaluated through simulations as discussed below.

Data from 1992–1996 were selected from the Heidelberg College data sets (Richards and Baker 1993) described in the Appendix. For the purposes of the statistical analysis in the Appendix, zero values were treated as missing observations; consequently, some stations had periods of missing data, mostly during late fall, winter, and early spring. Because concentrations of the three herbicides are typically low during these periods of missing values, the missing values were estimated so that there were roughly weekly observations but no more than about 2 weeks between observations. The period of missing records was broken into four periods: September, October–December, January–March, and April. The median concentration of the available data for each period was used to estimate the missing values in each period. Piecewise linear interpolation was then used to obtain 24 values per day for each day of the year. The resulting data sets (one for each herbicide) then served as true series for the subsequent simulation study described below.

The nonparametric approach was used to test the breakout group's best professional judgment of collecting 33 samples per season at one site by examining a data set for the herbicide alachlor from a creek in Ohio (Appendix). These data were used to form some true, daily time series, which were subsampled at 33 samples per season to obtain estimated percentiles. The estimates were compared with "true" values calculated from the entire original series. Table A2 in the Appendix shows three percentiles of the alachlor concentration distribution (in µg/L). The results suggest that the sample size recommendations are adequate. For example, the 95th percentile of the "true" distribution is 2.666 µg/L, whereas that estimated from the 33 samples is 3.594 µg/L, with lower and upper confidence limits of 1.39 and 5.97 µg/L, respectively. Thus, for this data set at least, the recommendation of 33 samples per season produced 50th, 95th, and 99th percentile estimates that were well within the 90% confidence interval and between 1% and 25 % of the "true " estimates.

However, because it was not clear how representative this particular data set was of the temporal pesticide distribution in surface water, a second set of simulations was carried out to investigate the effect of different sampling strategies and to see whether the conclusions from the previous analysis varied with station location, herbicide type, and relative amount of rainfall ("wet" years versus "dry" years). Four sampling strategies, chosen to reflect those used in practice (monthly, twice monthly, and four and 10 times a month), were evaluated for three herbicides at each of the four Ohio streams. For all scenarios, one observation was randomly selected from each sampling period of the time series. As before, percentiles were estimated from the sampled data and compared with those obtained from the "true" data sets, and this procedure was then repeated 50,000 times.

Tables A3–A8 in the Appendix provide the results for two streams, one representing the smallest drainage area (Rock Creek) and the other representing the largest drainage area (Maumee River). Results from the other two streams were similar and are not included. For instance, where an annual herbicide concentration distribution is of interest, results indicate that monthly and twice-monthly sampling may not be sufficient for accurate estimation of higher percentiles (90th, 95th, and 99th). Sampling four times a month or weekly is adequate for estimating the 50th, 90th, and 95th percentiles, but is not sufficient for estimating the 99th percentile. Sampling 10 times a month may be needed to estimate the 99th percentile, but such estimates may not be very accurate because the standard deviations associated with the 99th percentile estimates are fairly large. Moreover, the precision associated with the percentile estimates in "wet" years when concentrations are higher is much lower than the precision associated with percentile estimates in dry years when concentrations tend to be lower. In general, the nonparametric estimation procedure tended to overestimate higher percentiles, particularly for smaller numbers of samples (96 samples or less).

The results in the Appendix indicate that weekly sampling may be adequate for estimating the 95th percentile (but not the 99th percentile) of daily mean concentrations for acute assessments, which is in agreement with similar conclusions reached by the Research Triangle Institute (RTI) during the design of a national drinking water survey for the American Crop Protection Association (ACPA) and the EPA (ACPA 1999). RTI estimated that 59 random samples would be necessary for the 95th percentile, which increased to 298 for the 99th percentile. One important conclusion from this analysis is that the random sampling is from the total domain. If one wants to obtain the 95th percentile for each of three different seasons, 59 samples would be required in each season. However, the number of samples does not increase as the number of drinking water facilities increases, as long as the desired objective is a concentration distribution across all of the specified drinking water facilities (not distributions for each individual facility). Note that this is not the approach used in

Table 2.4 . Gross vulnerability factors to identify potential high- and low-risk locations for surface and ground water[a] supplies

Surface water		Ground water	
High risk	**Low risk**	**High risk**	**Low risk**
High use	Low use	High use	Low use
Close to application	Far from application	Close to application	Far from application
High runoff	Low runoff	Fast recharge	Slow recharge
Small impoundment	Large river	Shallow well	Deep well
Small contributing area	Large contributing area	Short travel time	Long travel time
Attach population to each			

[a]Not under the influence of surface water.

this report, where regional distributions are derived based on distributions for individual facilities. Because of the extra information on the individual facilities obtained using the approach described in Table 2.1, more samples are required to obtain regional distributions.

2.2.4. Sampling Locations

Constructing a representative pesticide distribution at the regional and national scales leaves many options on how to choose the surface water supplies for sampling. One preferred method is to sample candidate systems at random over the assessment area from a pool of appropriately stratified and weighted supplies. First, sampling should be focused on water supplies within assessment areas where the pesticide has been applied, i.e., the use region. Then the use region should be stratified by degrees of risk of having a pesticide migrate to a water supply. Finally, because it often will be desirable to get a more accurate estimate of concentrations at the upper percentiles of the distribution, some form of weighting of supplies should be performed so that a greater number of samples is collected from the higher-risk water supplies. Additional details on these steps are provided in the following paragraphs.

2.2.4.1. Stratification

The sampling population should be stratified by classifying the use regions and assessment areas into higher- and lower-risk isopleths and ranking individual supplies by vulnerability. These risk isopleths could be based on the combination or nexus of Geographic Information Systems (GIS) layers depicting various factors that affect pesticide migration risk, including intensity of pesticide use, climatic factors (rainfall history), contributing-area characteristics (geology, slope, etc.), source of supply, type of water delivery function, and proximity of the application area (e.g., see Table 2.4). These factors determine for a particular assessment scale which portions of supplies are to be considered in the higher-risk categories (to be sampled more intensely) and which fit the lower-risk categories. Although population should not be used as a criterion in this high-risk/low-risk stratification, each water supply should have a record of the population associated with it for use in the subsequent risk assessment.

The key to successful stratification is the correct classification of site characteristics and ranking of vulnerability criteria. These criteria may differ from assessment area to assessment area and among supplies. Barring direct experimental evidence, stratification criteria and relative rankings are often a matter of opinion that may change with experts. For example, high-use areas at risk according to a model that emphasizes only the intensity of use might include parts of eastern Colorado, where wheat is grown and pesticides are heavily applied. However, low precipitation and little or no runoff and return flow suggest that the risk of drinking water contamination might be less than in other areas of the United States. Therefore, for eastern Colorado, the high-use area may be classified as low risk owing to mitigating geological and climatic factors.

The presumption of vulnerability will also change with scales. At a local scale a particular sandy soil location could be considered "high risk," whereas at a regional scale it might be a specific type of watershed and at a national scale a geologic formation.

Hence there is no a priori guarantee that candidate supplies classified as vulnerable and located in the higher-risk zones will necessarily show high levels of residues, because there will always be a certain amount of uncertainty when the classification criteria are matched to conditions. However, in a statistically valid stratification approach, exposure estimates will still be accurate, although the efficiency of the sampling design will be degraded slightly and confidence in the estimate may be lower.

2.2.4.2. Weighting

In attempts to define a statistical probability distribution of the pesticide's concentration in the drinking water, it is essential to draw samples from every part of the distribution. Although most samples might be drawn from vulnerable supplies in the high-risk zone, a few samples will need to come from the drinking water supplies in the no-risk and low-risk areas and a few from the middle.

The following simplified example illustrates the concepts and pitfalls associated with stratification and weighting. Let the rectangle in Figure 2.2 represent the pesticide use region within an assessment area. Using appropriate GIS overlays of criteria important to pesticide migration in this area, one stratifies the use region into two areas: high risk (~90% of the area) and low risk (~10% of the area). (For simplicity, only two categories are proposed and one threshold; in fact, any convenient number [quartiles, quintiles, percentiles] could be used.) To obtain a high-accuracy distribution for this region, 20–30 water supplies should be chosen for sampling within this use region (see the second column of Table 2.2). One strategy would be to allocate the sample number based on the relative area in each stratum, for example, 10%, or three water supplies, from the low-risk area and 90%, or 27 water supplies, from the high-risk area. Unfortunately, this method would not be practical in the more usual case where 70% of the area could be considered low risk. In this case, sampling 21 of the 30 supplies (70% of the supplies) in the low-risk area would be wasteful of resources because of the high probability that the majority of, or even all, samples would result in nondetects.

An alternative way is to assign a certain percentage, for example, 10%, of supplies selected for monitoring to the low-risk area, regardless of the area size, and weight that 10% so that the weights would be inversely proportional to the probability of pesticide residues being found there. This would expand the data without the necessity of actually taking samples many of which would probably turn out to be nondetects. Returning to the original example of three water supplies in the low-risk zone, each would be weighted 30/3, with the remaining 27 supplies having the weight of 30/27. For the case of 70% of the area being in the low-risk zone, the three supplies would be weighted 30/21 and the remaining 27 would be weighted 30/9. The procedure is a form of data expansion by weighting. The percentages (10% in the low-risk zone and 90% in the high-risk zone) suggested here are arbitrary, and higher or lower values could be used.

If there were more than 30 candidate water supplies to choose from in a given assessment area, similar (homogeneous) locations should be grouped together for sampling purposes. Each one of the water supplies within an assessment area in a use region stratified by intensity of use would have a vulnerability ranking associated with it. Relative weighting of the sampling universe of water supply candidates based on intake characteristics with the gross vulnerability criteria in Table 2.4 could lead to relative ranking of water supplies into high- and low-risk zones, and suggest candidates that could be picked as the three low-risk or 27 high-risk water supply locations for sampling. If desired, the above approach could potentially also identify more vulnerable locations for additional sampling of supplies at the upper tail of the distribution.

This discussion assumes that water supplies selected for sampling are spatially independent. If significant autocorrelation among locations is present, additional samples may be needed to produce defensible probability distributions (Appendix).

Preexisting sampling designs, such as NAWQA (Gilliom et al. 1995, Gilliom and Thelin 1997), the Environmental Monitoring Assessment Program (EMAP) (Jones et al. 1997), and the National Pesticide Survey (NPS) (EPA 1990), can also be examined as possible guides to stratification and weighting in identifying candidate supplies for sampling. Although information from past pesticide studies (e.g., Larson et al. 1997) may be used to estimate probability distributions at a low accuracy level and to identify potential sampling locations in a gross exploratory manner, caution is advised in interpreting the findings, since some of the sites in those studies were selected for reasons not related to drinking water.

2.2.5. Ancillary Data

Depending on the scope and purpose of the monitoring study, ancillary information collected before sample collection can be used to stratify a use region or assessment

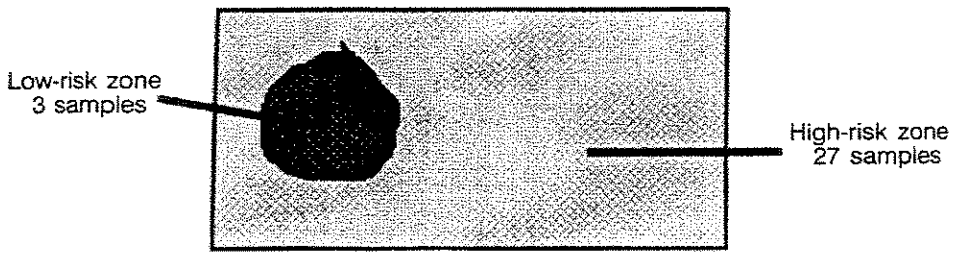

Figure 2.2. Example of a pesticide use region stratified into lower- and higher-risk zones based on region-specific vulnerability criteria.

area by vulnerability to indicate preferable sampling locations, to rank individual water supplies by their vulnerability criteria, or to tailor sampling strategy to local conditions. Ancillary information collected during and after monitoring can assist in the interpretation of results. Both types of information can be useful for model calibration and verification.

The ancillary data useful for site vulnerability assessment are of two different kinds: regional and local. The regional data sets available nationally include, for example, the USGS Digital Elevation Models (DEM) that describe landscape topography; others contain information on climate, soils, hydrogeology, and type of cover in the assessment area within a use region. This type of information is usually available in a GIS format at a large scale. Local conditions are more important in terms of water supply vulnerability to pesticide contamination. Information such as the amount and distribution of precipitation, timing of pesticide application, stage of crop growth at the time of application, size of the contributing area relative to the application area, type of source water (e.g., size of impoundment, stream, river, etc.), and distance of the application areas from the intake of the drinking water supply may all be important for data interpretation. This type of ancillary information is site specific and usually better known locally. Note, however, that the amount of ancillary data required to be collected will vary significantly between studies and that the collection of such data should be guided by the scope and purpose of the monitoring study and risk assessment.

2.3. Sampling Plans for Ground Water Supplies

A number of issues unique to ground water dictate that a monitoring strategy for ground water supplies should differ significantly from one derived for surface water supplies. For example, pesticide concentrations in ground water can vary greatly among wells in the same aquifer, but concentrations will typically change relatively slowly because of the slow rate of leaching and ground water flow. Because of slow temporal changes, ground water also typically exhibits much lower peak concentrations than surface waters.

For ground water not under the direct influence of surface water, the highest rate of recharge to the water table occurs during the dormant season. This does not imply that pesticide concentrations in drinking water wells will necessarily be highest during the dormant period, because additional travel time will usually be needed for pesticide residues to reach a well. Conversely, contamination of the water table from point sources, accidental spills, poor well construction, back siphoning, and faulty well casings can occur at any time, although it is more likely during the height of the pesticide application season. The agricultural industry has long recognized this potential for accidental contamination of drinking water supplies and has undertaken successful voluntary efforts to educate pesticide users on proper handling practices (D. Gustafson, Monsanto Company, personal communication, 1998).

Regardless, peak concentrations in ground water generally will not be as pronounced as they are in surface water and will be spread out over a much longer period of time. Therefore, given that daily peak concentrations are approximately equivalent to the annual average, ground water sampling in general should be spatially more extensive, but the frequency of sampling can be much lower than in surface waters.

Additionally, use differs between surface and ground water supplies. Typically, many people draw from a single point in surface water systems, whereas fewer people draw from several points on an aquifer. Therefore, ground water monitoring designs may need to be tailored to consider the use of the resource. For example, if the ground water supply is a community system similar to the typical surface water situation where consumers draw from one source, temporal sampling may be desirable to generate a time series and frequency distribution unique to that system. Conversely, in rural assessment areas, where there are very small populations scattered over wide areas and many individual domestic wells, a statistical design that uses a more spatially extensive sampling strategy should be considered.

Furthermore, to represent long-term temporal variability in the ground water, the sampling plan will need to ensure that the pesticide residues have had adequate time to reach the aquifer and associated wells. Research indicates that pesticide residues may take up to 5 years, and sometimes longer, to migrate from the surface to the ground water (Barbash and Resek 1996). Ideally, the population of sites should be stratified by the projected travel and arrival time at the well to capture the temporal distribution of pesticides in the ground water. If relatively complete fate and transport information is available, stratification of the projected travel and arrival time at the well may be possible for compounds that have been registered for 5 or 10 years, but poses a serious obstacle for assessing new chemicals. Mathematical models may be particularly helpful in predicting time of arrival and duration.

A final difficulty is that for rural wells in particular, rarely is there complete knowledge of basic information such as what strata a well is drawing water from, the direction of flow in the aquifer, travel time, overburden properties, size and extent of percolate contributing area, and whether the well is hydrologically connected to the pesticide application field or not. In fact, many of the details

necessary for appropriate stratification and weighting of individual water supplies may be either missing or uncertain.

2.3.1. Sampling Plan to Estimate Peak Concentrations (as Daily Means) and Long-Term Mean Concentrations (as Annual Time-weighted Averages)

The matrix in Table 2.5 indicates a best professional judgment of the suggested number of samples needed over time to characterize the temporal distribution of pesticide residues in an individual drinking water supply. The matrix also describes a minimum number of supplies required per assessment area to develop statistical probability distributions at a regional and national scale for each level of accuracy. Discussion of the matrix is followed by a statistical evaluation of the sampling plan at a regional scale and suggested guidelines on how to choose sampling locations.

2.3.1.1. Sampling Plan for a Local-Scale High-Accuracy Assessment

A local-scale assessment implies some type of site-specific assessment of an individual water supply, typically designated "most vulnerable." As discussed earlier, defining a priori any one supply as the most vulnerable in a use region is highly problematic and probably the least likely scenario to be assessed. Stratification and weighting by site vulnerability criteria do not always produce the desired result (EPA 1990). Well sites that test positive for residues during an initial sampling of a use region within an assessment area might be considered vulnerable. However, it is seldom known what percentage of drinking water supplies such vulnerable wells represent, since this varies with the sensitivity of the test. Therefore, for risk assessments that need to define the upper tail exclusively or particularly accurately, sampling of a class of vulnerable supplies is a more logical approach. However, for cases when an assessment of an individual ground water supply is desirable, guidance is provided in Table 2.5 on the number of samples that should be considered for collection.

Similar to the sampling plan for surface water, the requirements at the high-accuracy level for a local ground water supply represent a balance between what is realistically feasible and the minimum necessary to describe the temporal probability distribution of pesticide concentration in the drinking water. The local ground water supply identified as highly vulnerable might be a domestic well containing some residues located in a heavy agricultural pesticide use area close to the application site. Although such rural wells represent only a small part of the population of drinking water supplies, they represent the highest risk. Because peak and average concentrations are approximately equivalent for ground water, it was suggested that four samples per year be collected to fully characterize the temporal distribution.

As already noted, the exact length of the sampling cycle will depend on site characteristics and climatic conditions that influence the projected travel and arrival times

Table 2.5. Proposed sampling matrix for daily mean (peak) and/or annual time-weighted average pesticide concentration in ground water[1] supply wells at three spatial assessment scales and three levels of accuracy

	Spatial assessment scale		
Prediction accuracy level	**National assessment (multiple regions with multiple water supplies)**	**Regional assessment area (multiple water supplies)**	**Local assessment (single water supply)**
High	50 water supply wells per region 5 municipal, rest rural sampled once per year aggregated over regions	200 water supply wells per assessment area 20 municipal, rest rural sampled once per year	Single water supply well sampled 4 times per year
Medium	25 water supply wells 3 municipal, rest rural sampled once a year aggregated over regions + assumptions about the distribution	50 water supply wells 5 municipal, rest rural sampled once per year + assumptions about the distribution	1 water supply well sampled once every 3–5 years + assumptions about the distribution
Low	Existing data[b] from other wells in each region aggregated nationally + many assumptions about the distribution	Existing data from other wells in the assessment area + many assumptions about the distribution	Existing data from nearby wells + many assumptions about the distribution

[a]Ground water not under the influence of surface water.
[b]Public water supply wells are required to take four samples and four replicates per year; no such requirement is imposed on private rural wells.

of pesticide residues at the well. Sampling should be concentrated over this period, with a few samples before and after to make sure that the predicted arrival time and duration are correct. To characterize temporal variability in the ground water (not under the influence of surface water), known vulnerable supplies will need to be sampled over time close to the application area.

2.3.1.2. Sampling Plan for a Regional-Scale High-Accuracy Assessment

To be consistent with the surface water approach, a regional ground water supply assessment area was envisioned as a collection area comprising all individual and municipal water supply wells that draw water from different aquifers within an agricultural use region, such as the corn belt. Alternatively, it could comprise individual and municipal wells drawing water from one particular (geologically distinct) confined aquifer system within a pesticide use region. At the regional scale, it is reasonable to consider monitoring 200 wells in the pesticide use region of each assessment area, with a certain percentage (e.g., 10%) of the samples drawn from municipal supplies and the remainder (90%) from rural wells (Table 2.5). These percentages are arbitrary and should be adjusted depending on the goals of the assessment. The rationale for this split is that although rural domestic wells represent the highest risk and outnumber community ground water supply systems by ~100:1 (EPA 1990), the population served by the community systems is larger than the rural population by more than 2:1 (Solley et al. 1993).

2.3.1.3. Sampling Plan for a National-Scale High-Accuracy Assessment

The national ground water assessment universe was considered to be the collection of several regional ground water assessment areas. To obtain a high-accuracy assessment at the national scale, at least 50 ground water supplies per assessment area should be sampled once per year, with perhaps five wells (10%) located in municipal supplies. Table 2.6 shows the total number of samples needed at the local scale in time and at the regional and national scales in space. Note that the national total is estimated based on five regional assessment areas. When ground water is considered, single-year results are seldom informative enough, and monitoring at both the regional and national scales may need to be conducted for more than 1 year or every few years. However, for products that have been used for many years, additional sampling efforts might best be put into determining spatial variability rather than temporal variability. Temporal sampling becomes more critical if an attempt is being made to determine trends as a result of specific actions, such as implementation of management practices or other changes in pesticide use.

Table 2.6. Number of samples recommended for estimation of the daily mean (peak) and/or annual time-weighted average concentration of pesticides in ground water supplies at three spatial assessment scales and two levels of accuracy

Accuracy Level	National[a]	Regional	Local
High	250	200	4
Medium	125	50	0.3[b]

[a] Assume five assessment areas nationally.
[b] Sampled once every 3 years.

2.3.1.4. Sampling Plan for Local-, Regional-, and National-Scale Medium- and Low-Accuracy Assessments

The breakout group recommended that, similar to surface water, medium and low relative accuracy levels of assessment for ground water could be achieved either with fewer samples or by relying on existing data plus assumptions. Although this approach is incorporated in the Table 2.5 matrix, the recommended numbers of samples are considerably fewer and the distributions are expressed on an annual or multi-annual basis, because the response time for the ground water is expected to be longer than for the surface water. However, because of lower concentrations (approximately one order of magnitude), implementation of lower detection levels is advocated. Sampling requirements are reduced in assessments at medium and low relative accuracy levels by substituting statistical assumptions about the distribution and/or substituting existing data for actual measurements. If the existing data appear relevant and represent the only information available, they could be applied to the specific situation when supplemented by increased reliance on statistical assumptions about the distribution.

In the sampling matrix at a low level of accuracy, existing databases on ambient, background ground water quality could be used. However, it is recognized that these historical databases do not necessarily represent drinking water supplies or conditions in the drinking water as used by the public and private (rural) water supply systems. For example, the USGS NAWQA (Gilliom et al. 1995) database is stratified by urban wells, shallow wells, and deeper wells that are not necessarily representative of drinking water exposure to the population, and may include shallow ground water under the influence of surface water. The data may be sufficient to indicate that concentrations are below levels of concern. If results are questionable, additional analyses may be warranted at the

actual drinking water supply sites. The other potential problem associated with the historical data sets is the inability to estimate the travel time of pesticide residues, because rarely if ever is there a record of sufficient duration to describe a complete travel time cycle for a specific location (Barbash and Resek 1996). This is particularly critical for ground water supplies where pesticide residues might not yet have arrived at the water table.

2.3.2. Statistical Evaluation

The numbers of samples and numbers of water supplies to sample in Tables 2.5 and 2.6 were derived using best professional judgment tempered by considerations of practicality and cost. To provide some perspective on these numbers, a statistical analysis was performed at a regional scale to indicate what the breakout group's estimate on sample numbers produced in terms of probability (e.g., 0.75, 0.90, 0.95) and percentiles (50th, 90th, 95th, 99th) (Appendix). The scope of this statistical analysis for ground water is limited to the question of the adequacy of the number of water supplies to sample in a use region. Data used in this example were atrazine concentrations obtained from sampling 770 ground water wells in Nebraska. These were subsampled randomly at 30, 50, 100, 200, 300, and 400 wells, and a simulation approach used previously for the surface water was implemented.

Tables A9 and A10 in the Appendix suggest that the recommendation of sampling 200 wells per use region may be adequate. The 95th percentile of 0.12 µg/L is within 8% of the true value and contained within a relatively narrow 90% confidence interval (0.00–0.31 µg/L); an increase in sample size to 300 and 400 does not greatly improve the accuracy. Estimation of the 99th percentile requires more samples (400), but again the increase in precision (i.e., the spread of the confidence intervals) between 200 and 400 is small.

Potential impacts of spatial scale were investigated by increasing the spacing between samples while maintaining reasonable coverage (Appendix). Table A10 in the Appendix suggests that although the sample size of about 200 is adequate for the estimation of the 90th percentile, for the 95th percentile the precision improves substantially by doubling that number.

Computations in the Appendix assume independence for both the surface water and the ground water. However, concentrations of pesticide residues are likely to be correlated both in space and in time. This would create an additional burden in terms of developing the appropriate sampling strategy and would change the number of samples required. Should autocorrelation be present, the required sample numbers may need to be approximately doubled (Matalas and Langbein 1962).

2.3.3. Sampling Locations

Several considerations unique to ground water make the development of complete spatial and temporal probability distributions at the regional and national scales particularly challenging. These considerations include the large number of individual domestic supplies, variable time of travel, a lack of site-specific geologic and hydraulic information, and an occurrence distribution that is often heavily skewed toward wells in close proximity to heavy pesticide application areas. Therefore, in contrast to the surface water strategy, the sampling strategy for ground water is designed primarily to identify areas of potential concern.

A two-stage type of sampling is proposed. First, to create a statistical probability distribution, there is a need to sample water supply wells everywhere within the assessment area with a known, finite probability of being chosen. Second, if there is a need to better assess vulnerable supplies, a more detailed sampling would be required of locations where the residues are detected in the first stage.

In the first stage—creation of a statistical probability distribution—because all water supplies are part of the sampling universe, each supply initially must have an equal chance of being chosen irrespective of adjacent land use and location relative to the application field. However, practical monitoring experience indicates that pesticides, if detectable at all, are most likely to be detected in monitoring wells located within half a mile of a treated site (EPA 1990; R. Jones, personal communication, 1998). To maximize the possibility of detection, ground water sampling therefore needs to focus on two types of information that are usually available: intensity of use in the surrounding area and proximity of the water supply to the application site. Water supply wells everywhere within the assessment area should be stratified by some general vulnerability criteria into high- and low-risk zones (Table 2.5) and heavily weighted toward sampling sites close to heavy application areas (e.g., within 1000 feet). By focusing primarily on worst-case areas, this first-stage ground water–related sampling would indicate whether there is concern with a specific pesticide in a given assessment area. Potentially it would also identify some high-risk, vulnerable supplies. The use of existing sampling designs such as those of the NAWQA program (Barbash and Resek 1996, Gilliom et al. 1995, Gilliom and Thelin 1997), of the National Pesticide Survey (EPA1990), and of Gustafson (1993) may also provide guidance in the identification of concentrations and candidate ground water sources for sampling in a gross exploratory manner and provide insights into mechanisms controlling pesticide transport in the area. For example, atrazine has now reached the ground water in selected locations (EPA 1990, Barbash and Resek 1996).

Information about the spatial and temporal occurrence of atrazine and other pesticides that have been in use for 5 or 10 years may provide insight into local modes of travel and help identify potential travel pathways.

Although GIS-based regional-scale stratification might be very useful in estimating the isopleths of higher and lower risk, the results should be used with caution. Predictions may not be accurate locally, since some properties, such as pesticide travel and arrival times, depend on local conditions. In the sampling of rural domestic wells, rarely is complete basic information available such as which strata the well is drawing water from, the direction of flow in the aquifer, travel time, overburden properties, the size and extent of the percolate contributing area, or whether the well is hydrologically connected to the pesticide application field or not. In fact, many of the details necessary for the appropriate stratification and weighting of the individual water supplies may be either missing or uncertain. The number and location of domestic wells in a given county may not even be known accurately. This provides a rationale for systematic sampling at different spatial and temporal scales of locations known or presumed to be vulnerable.

In the second stage—where there is a need to evaluate location-specific risk or to better characterize vulnerable water supplies—a more detailed sampling of locations where the residues show up in the first stage may then be required. Although rural wells represent only a small part of the population of drinking water supplies, they typically are at the highest risk. A vulnerable supply might be a domestic well in an agricultural pesticide use region where water is drawn from a high-recharge aquifer, one that is overlain by sandy soils or otherwise permeable soils or rock types that allow rapid movement of water from the land surface to the aquifer. Qualitative descriptive information about the site and well as well as technical information about the area can be used to indicate preferable sampling locations and to rank the individual water supplies by their vulnerability criteria. Such an evaluation should consist of a detailed description of local geology, soils, and site location relative to landscape configuration (concave versus convex, foot slope versus head slope). The information gathered may be evaluated for potential vulnerability and compared with a similar type of stratification elsewhere.

However, even with perfect understanding of an assessment area's characteristics, which are seldom available, it is still not certain, barring extended sampling in time, where the vulnerable wells are, which application fields they are connected to, and how long it will take for the compound to reach the ground water.

It is unclear how many wells are located, for example, within 1000 feet and beyond 1000 feet of the application area. It is recognized that such sampling is severely skewed toward the worst-case scenario, and at best describes only the upper tail of a probability distribution. However, because pesticide distributions in ground water tend to be very uneven (EPA 1990), residues are not likely to be found on a regular basis even within 1000 feet of some application sites. When a number of ground water samples within a heavy-use assessment area exhibit concentrations below levels of significance in aggregate risk assessment in close proximity (within 1000 feet) to the application site, this is probably an indication (barring a particularly long travel time) that pesticide contamination is not likely to be a problem. However, when a few water supply wells (out of a large number sampled) test positive, it may be difficult to construct a meaningful distribution, and may indicate anomalous site-specific behavior.

An additional concern with this sampling design is the problematic nature of matching an exposure assessment from ground water based largely on a worst-case sampling design with one from surface water based on statistical probability. At a minimum, the discrepancy would have to be acknowledged and appropriate caveats added. One mitigating factor is that risk assessments of exposure are usually done separately for ground water and surface water supplies, and respective concentrations will rarely be equal.

2.3.4. Ancillary Data

Depending on the scope and purpose of the monitoring study, ancillary information collected before sample collection can be used to stratify a use region or assessment area by vulnerability, to indicate preferable sampling locations, to rank individual water supplies by their vulnerability criteria, or to tailor a sampling strategy to local conditions. Ancillary information collected during and after monitoring can assist in the interpretation of results. Both types of information can be useful for model calibration and verification. However, the amount of ancillary data that must be collected will vary significantly between studies and should be guided by the scope and purpose of the monitoring study and risk assessment.

The ancillary data useful for the ground water vulnerability assessment are different from those associated with surface water supplies. Because there is little direct information regarding the structure and behavior of the soil–vadose-zone–aquifer system, most spatial and temporal properties need to be inferred from information usually obtained elsewhere. Consequently, the role of ancillary data is far more critical in predicting the migration of pesticides to ground water than surface water. Although the emphasis for the latter was on runoff with a relatively short travel time to the stream, the emphasis in ground water is on the distribution of percolation flux, extended

travel time, and connectivity of the preferential flow pathways. The movement of the wetting fronts may in general be too slow to provide a viable pathway to the ground water within a reasonable time frame. Consequently, the type of ancillary information that is often most urgently needed is the characterization of the preferential flow pathways that provide a rapid conduit to the water table. Data sets that contain information on hydrogeology of the vadose zone as well as the chemical and physical characteristics of soils, especially spatial distribution of hydraulic conductivity and macroporosity in the use region, may contain clues to the distribution of recharge flux over an area and in time (Barbash and Resek 1996). This type of information may reflect aquifer vulnerability to pollution from pesticide residues and is often available in GIS format at the local, regional, and national scales.

2.4. Sampling Plans for Ground Water Supplies Under the Influence of Surface Water

Shallow wells, ground water overlain by highly permeable materials, wells where withdrawals are compensated by inflows from rivers, karst terrain, and in general any ground water under the rapid influence of surface water should be considered separately from wells deriving water from deep confined aquifers. The sampling matrix applicable to these situations would follow that proposed for the surface water for both the daily mean (Table 2.1) and the long-term average mean (Table 2.2), with the stipulation that sampling sites should be stratified and weighted such that most are selected in close proximity (e.g., within 1000 feet) of the source of contaminated recharge or the application field.

The response of the ground water under the influence of surface water is not well understood. Although much faster than the recharge to the ground water, this response is usually somewhat delayed and less pronounced than that of a stream because of the travel time associated with the movement of pesticides through the soil and vadose zone in the network of cracks, fissures, and macropores. The well response varies from region to region and requires good local knowledge of the sampling environment and tailoring of the sampling strategy to site-specific criteria. Shallow rural domestic water supply wells are generally the ones most likely to be affected. These wells, because of their sheer numbers and the lack of documented knowledge about well characteristics and construction, are also least likely to be found, identified, and remedied. Pesticide concentrations are likely to be highest in these wells.

2.5. Summary and Conclusions

The authors of this chapter were charged with developing a monitoring strategy to ensure that data on pesticide residues in drinking water are collected such that they are compatible with other exposure data collected and used in a probabilistic aggregate exposure analysis. These questions were addressed by providing guidance on the development of a space-time sampling design to monitor pesticide concentrations at the national, regional, and local scales. To characterize the upper tail of the distribution of pesticide residues in drinking water, sampling needs to focus on vulnerable locations, high-risk site characteristics, and sampling times when pesticide residues are most likely to be present. In contrast, the sampling strategy to describe a probability distribution requires that all water supplies and times have a known finite chance of being sampled. Appropriate stratification and weighting can be used to resolve apparent contradictions between the two approaches. It should be emphasized that even when only the high-risk vulnerable locations are targeted, a limited number of supplies may have detectable residues, making it difficult to construct a probability distribution at a high accuracy level. Under these circumstances, medium and low levels of accuracy predictions may be the only ones that can be obtained.

This chapter presents preliminary ideas in support of this design task and suggests some guidelines regarding sampling requirements applicable at the local, regional, and national levels with a high, medium, and low degree of accuracy. The recommendations are based on the best professional judgment of the breakout group, supplemented by illustrative examples from field data. The recommendations center primarily on the type, size and nature of the data sets needed. Guidance on how and where to choose sampling locations is provided. Although surface water and ground water are discussed separately, guidelines developed for surface water are assumed to apply to ground water supplies under the influence of surface water. Additional consideration is given to variability and uncertainty as related to the size of assessment area, to factors affecting vulnerability of water supplies, and to ancillary data needs.

For surface water supplies, three matrices of scale versus degree of accuracy were designed, and the number of samples required for each assessment area and level of accuracy over time was identified, while addressing both daily (peak) and annual (chronic) probability distributions. At a local scale, the proposed sampling strategy provides sampling guidelines for detection of pesticide residues in a specific water supply. Because identifying a priori which supply is the most vulnerable one is very problematic, it is more likely that monitoring to characterize vulnerable supplies will encompass a *class of water supplies* thought to be vulnerable because of known site characteristics. A regional assessment area was envisioned

as the aggregation of several water supplies within a watershed or across watersheds within a pesticide use region, whereas the national-scale proposed distributions were to be aggregated by geographic regions. To show the effects of treatment, it is recommended, if feasible, that both intake and tap samples be collected and that the tap samples be analyzed only if the intake samples exhibit detectable concentrations. Monitoring recommendations were tempered by the high cost of analysis, practicality, and applicability of the proposed strategy.

To develop a probability distribution, it is important to characterize the whole distribution at each assessment scale. The expected level of accuracy in the monitoring matrix is to be based on the quality and quantity of available data. Increased levels of accuracy require more and higher-quality data and fewer assumptions. For example, sampling surface water at a high level of accuracy for peak daily means may require at least 33 samples per season at the local level, 20–30 water supplies per assessment area within a use region at the regional level, and 10–15 water supplies per geographic region at the national level. Decreased levels of accuracy would require fewer data, but must be supplemented by statistical assumptions and/or existing measurements. Similarly, the numbers proposed for the annual (chronic) matrix would be 12–24 samples collected monthly or twice monthly at the local level, with the same numbers of water supplies as detailed above at the regional and national levels. The number of supplies and samples identified at each level of accuracy and assessment scale combination in the accompanying tables represents the minimum necessary to develop a statistical probability distribution. If sampling were conducted for the purpose of characterizing the upper tail only, fewer samples and supplies may be sufficient.

Systematic sampling with a random starting time for the surface water supplies, with more samples taken during the pesticide application season, was a preferred approach, since it helps ensure representative samples and can make use of statistical theory regarding the estimation and precision of estimates.

The recommended approach for choosing which supplies to sample was first to stratify the population, using appropriate vulnerability criteria, into low- and high-risk zones. For a regional assessment, the recommendation was to draw a certain percentage of samples (e.g., 10%) from the low-risk area, regardless of size, and the remainder (e.g., 90%) from the high-risk area, with each sample inversely weighted by the number of samples drawn, with more samples taken during the application season. This approach is a form of data expansion. To construct a credible probability distribution, it is particularly important to draw samples (appropriately weighted) from all expected levels of vulnerability and from the whole population of supplies within the pesticide use area. Weighting and stratification were considered acceptable alternatives in order to target most samples to drinking water supplies in the high-use most-vulnerable zones serving the largest population.

A different monitoring strategy was devised for ground water because of issues unique to ground water. To develop a credible national or regional pesticide distribution in ground water supplies, a two-stage type of sampling approach was recommended. First, to create a statistical probability distribution, there is a need to sample water supply wells everywhere within the assessment area with a known, finite probability of being chosen. The assessment area should be stratified by some general vulnerability criteria into the high- and low-risk zones. In the second stage, to evaluate location-specific risk and for detection purposes only, a more detailed sampling of locations where the residues show up in the first stage would then be required. Under these assumptions, it was recommended that at a local-scale high-accuracy level, a single vulnerable well be sampled four times per year, with 200 wells per use region at a regional scale and 50 wells per geographic region at a national scale. Decreased levels of accuracy would again require fewer data, supplemented by statistical assumptions and/or existing measurements. In locating sampling sites for ground water, the breakout group envisioned targeting site vulnerability criteria and high-risk zones using appropriate weighting and stratification approaches discussed in connection with surface water supplies. Compared with surface water runoff travel to streams, pesticide residues will take more time to arrive at the water table. Hence, the viable sampling cycle for ground water should be targeted at when pesticide residues are expected to show up in ground water. An adequate number of samples should be taken before and after this time period to ensure that the prediction is correct.

Depending on the scope and purpose of the monitoring study, ancillary information collected before sample collection can be used to stratify a use region or assessment area by vulnerability, to indicate preferable sampling locations, to rank the individual water supplies by their vulnerability criteria, or to tailor a sampling strategy to local conditions. Ancillary information collected during and after monitoring can assist in the interpretation of results. Both types of information can be useful for model calibration and verification. However, the amount of ancillary data required to be collected will vary significantly between studies, and should be guided by the scope and purpose of the monitoring study and risk assessment.

For both the surface and ground water monitoring scenarios, ancillary data on soils, hydrogeology, land use, and management are useful to identify potentially vulner-

able locations and plan the sampling strategy. Concomitant climatic data of sufficiently long duration associated with a specific location can help identify how representative the sampling season was. Although the report addresses the sampling plan for surface water supplies, ground water supplies, and ground water under the influence of surface water supplies separately, any given assessment area may incorporate all three pathways at the same time. Consequently, it may be necessary at times to collect sufficient ancillary information to characterize the most critical distribution pathways of the potential exposure in a proposed assessment area before initiating monitoring.

To provide some perspective on these strategies, a statistical analysis was performed to evaluate the best-professional-judgment recommendations in terms of confidence in estimating specific percentiles of the pesticide concentration distributions. These are presented in the Appendix to this chapter. The scope of analysis was limited to the question of the adequacy of the number of samples to collect at a site. The analyses considered only the scenarios in the high/vulnerable cell in the daily surface water sampling matrix and the high/regional cell in the ground water sampling matrix. The Appendix gives a very basic overview of some of the statistical approaches to the estimation of percentiles and highlights some of the issues involved when using these methods. A preliminary analysis of some "typical" data on pesticide concentrations in both surface water and ground water indicated that the sampling strategy and sample size recommendations based on the best professional judgment of the breakout group are minimally adequate at both the local surface water scale and the regional ground water scale. However, very accurate estimation of extreme upper percentiles (such as the 99th) may require a very large number of samples or a radically different sampling strategy. Weekly sampling is adequate for estimating the 50th, 90th, and 95th percentiles. Moreover, the precision associated with the percentile estimates in "wet" years where concentrations are higher is much lower than the precision associated with percentile estimates in dry years where concentrations tend to be lower. These estimates are for characterizing single sites and have the potential for some reduction in sample number with proper stratification based on the timing of the occurrence of residues. For studies in which multiple sites are being investigated, the number of samples collected per site is expected to decrease, perhaps allowing sampling every other week during the period when residues are expected to be present in significant concentrations, and at lower frequencies during periods when significant amounts of residues are not expected to be present.

Several statistical issues have arisen in the context of this report. First, what does the literature show on sample size recommendations based on the goal of estimating percentiles with a known confidence using nonparametric confidence limits? Second, what is the relationship between sample size recommendations from the tolerance-level approach and the accuracy of percentile estimates? Third, what are the statistical properties of the nonparametric percentile estimators and associated nonparametric confidence intervals? Are the coverage probabilities of these intervals accurate? How can nonparametric confidence limits be constructed for complex sampling strategies? What is the effect of autocorrelation (both spatial and temporal) on percentile estimates and confidence bounds? Finally, and perhaps most importantly, how can the correlations between the concentrations of different pesticides (if present) be used to the best advantage?

These issues are important ones that must be addressed to better define a realistic approach to exposure assessments based on pesticide concentrations in drinking water. The supporting studies presented in the Appendix, in addition to illuminating some of these issues and illustrating them with real data, provide an overview of several approaches to sample size determination for estimating percentiles of pesticide concentration distribution in drinking water and an objective assessment of this report's recommendations.

References

American Crop Protection Association. 1999. A national survey of pesticides in drinking water obtained from surface sources. Washington, DC: ACPA

Barbash JE, Resek EA. 1996. Pesticides in ground water. Chelsea, MI: Ann Arbor Press, Inc

Food Quality Protection Act. 1996. Public Law No. 104-170

Gilliom RJ, Alley WM, Gurtz ME.1995. Design of the National Water-Quality Assessment program: occurrence and distribution of water-quality conditions: US Geological Survey Circular 1112. Washington, DC: USGS

Gilliom RJ, Thelin GP. 1997. Classification and mapping of agricultural land for national water quality assessment. US Geological Survey Circular 1131. Washington, DC: USGS

Gustafson DI. 1993. Pesticides in drinking water. New York: Van Nostrand Reinhold

Jones KB, Riitters KH, Wickham JD, et al. 1997. An ecological assessment of the United States Mid-Atlantic Region. EPA/600/R-97/130. Washington, DC: US Environmental Protection Agency

Larson SJ, Capel PD, Majewski MS. 1997. Pesticides in surface waters. Chelsea, MI: Ann Arbor Press, Inc

Matalas NC, Langbein WB. 1962. The information content of the mean. *Journal of Geophysical Research* 67:3441–3448

Olsen AR, Sedransk J, Edwards D, et al. 1999. Statistical issues for monitoring ecological and natural re-

sources in the United States. *Environmental Monitoring and Assessment* 54:1–45
Richards RP, Baker DB.1993. Pesticide concentration patterns in agricultural drainage networks in the Lake Erie Basin. *Environmental Toxicology and Chemistry* 12:13–26
Saaty TL, Alexander J. 1981.Thinking with models. New York: Pergamon Press
Solley WB, Pierce RR, Perlman HA.1993. Estimated use of water in the United States in 1990. US Geological Survey Circular 1081. Washington, DC: USGS
US Department of Agriculture. 1989. The second RCA appraisal: soil, water, and related resources on nonfederal land in the United States: analysis of conditions and trends. Washington, DC: US Government Printing Office
US Environmental Protection Agency. 1990. National Pesticide Survey: summary results of EPA's National Survey of Pesticides in Drinking Water Wells. EPA570990NPS5. Washington, DC: EPA Office of Water, Office of Pesticides and Toxic Substances

CHAPTER 2 APPENDIX. SAMPLE SIZE CONSIDERATIONS AND CALCULATIONS FOR ESTIMATING PERCENTILES OF DISTRIBUTIONS OF PESTICIDE CONCENTRATIONS IN DRINKING WATER

Carol A. Gotway-Crawford, National Center for Environmental Health, CDC

Charles G. Crawford, U.S. Geological Survey

This Appendix gives a brief overview of a nonparametric approach to estimating percentiles and evaluates several sampling strategies for estimating percentiles of distributions of pesticide concentrations in drinking water. There are also parametric approaches to percentile estimation and sample size determination. Although these are not discussed here, they should be used whenever possible, because if the parametric form of the distribution can be specified with reasonable confidence, parametric approaches are more accurate and more powerful. However, in practice, the parametric form usually cannot be specified with certainty, so this report focuses only on nonparametric methods.

Estimating Quantiles and Percentiles

Definitions

The pth quantile (also called a fractile) of a distribution of values is that value, X_p, for which a proportion, p, of the values is smaller than X_p. If F is the underlying cumulative distribution function, then $F(X_p)=p$. Quantiles expressed as percentages are called percentiles.

Given a desired proportion, p, consider estimating the pth quantile, X_p, from a set of data values (ordered smallest to largest), denoted as $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$. One approach is to estimate the pth quantile as the corresponding sample value $x_{(j)}$ where $j=np$. If np is not a whole number, conventions include rounding to the nearest sample value or interpolating between to the two closest sample values. This will give $x_{(n)}$ as the estimate of the maximum value, $X_{1.0}$. Because the true maximum is almost certain to be larger than the maximum value in the sample, this approach will give biased estimates. Thus, what is typically done is to estimate the pth quantile as the kth largest value in the data set where $k=p(n+1)$. If k is not an integer, the estimator of the pth percentile, denoted as $\hat{X}_p$, is then obtained by rounding down to the nearest integer or using linear interpolation between the two closest-ordered sample values. This is the approach that will be followed throughout the remainder of this report, with rounding instead of interpolation used to facilitate calculations.

Note: To use this method, the data set must be large enough so that for a given value of p, $k=p(n+1)<n$. Thus, to estimate the 95th percentile, at least 20 observations are needed, and to estimate the 99th percentile, at least 100 observations are needed. This alone gives a lower bound for required sample size.

Confidence Intervals

Confidence intervals for quantiles estimated using the methods described above can be determined in the following way. First, compute the rank values for the upper and lower confidence limits as

$$RLCL = p(n+1) - z_{\alpha/2}[np(1-p)]^{1/2}$$
$$RUCL = p(n+1) + z_{\alpha/2}[np(1-p)]^{1/2}. \quad (1)$$

These limits give the ranks of the sample values that form the confidence limits.

To illustrate this nonparametric approach to the estimation of percentiles and corresponding confidence limits, consider the following example. Eighty-two alachlor concentrations were collected in 1990 at Honey Creek station near Tiffen, Ohio (data courtesy of Charles Crawford, U.S. Geological Survey, from Peter Richards at Heidelberg College). Suppose we want to estimate the 90th percentile

with 95% confidence. The last 15 observations of the series (in µg/L) after sorting from smallest to largest are: $x_{(68)}$=3.331, $x_{(69)}$=3.362, $x_{(70)}$=3.387, $x_{(71)}$=3.643, $x_{(72)}$=5.030, $x_{(73)}$=5.291, $x_{(74)}$=5.332, $x_{(75)}$=5.953, $x_{(76)}$=6.328, $x_{(77)}$=7.768, $x_{(78)}$=11.241, $x_{(79)}$=11.606, $x_{(80)}$=11.813, $x_{(81)}$=13.786, and $x_{(82)}$=17.401.

Then $\hat{X}_{0.90}$ is the 0.90(82+1)=74.7th observation. Rounding down to 74 gives $\hat{X}_{0.90}=x_{(74)}$=5.332. Then $RLCL$=0.90(83)-(1.96)[(82)(0.90)(0.10)]$^{1/2}$=69.38, and similarly, $RUCL$=80.025. Then the interval defined by $x_{(69)}$=3.362 and $x_{(81)}$=13.786 gives a 95% confidence interval for $\hat{X}_{0.90}$.

Sample Size Determination

Determination of sample sizes using the nonparametric approach described above is not so straightforward, since the confidence limits are based on ranks. Thus, sample size considerations will be investigated using simulation as described elsewhere in this report.

Tolerance Intervals

One way to gauge the representativeness of a sample of values is by the extent to which the sample spans the distribution of values of the population. For example, if the sample contains both the smallest value and the largest value in the population, the sample spans 100% of the distribution of the population. It is possible to predetermine a sample size that will, with a specified probability, cover a desired percentage of the distribution of population values. This coverage is called a tolerance interval.

One-sided Tolerance Intervals

Suppose that high values of a variable are of interest and that the goal is to ensure with probability (1-α) that the maximum value for this variable in the sample will exceed 100p% of the population. In general, α will be a small fraction (e.g., 0.05) and p a large fraction (e.g., 0.95). This means that the maximum value in the sample will have a probability of (1-α) of exceeding the pth percentile of the population distribution, a desirable quality if estimation of percentiles higher than the pth percentile is desired. Table A1 gives the smallest sample size needed to ensure that the largest value in the sample exceeds at least 100p% of the population values. For example, taking p=0.95 and α=0.10, the maximum value in a sample of size 45 will have at least a 90% probability of exceeding 95% of the values in the population, or, said differently, the maximum value in a sample of size 45 will have at least a 90% probability of exceeding the 95th percentile of the population distribution. This is based on the equation (Hahn and Meeker 1991)

$$1-\alpha=1-p^n. \quad (2)$$

Table A1 gives corresponding sample sizes for a variety of confidence levels and desired proportions, and these are also presented graphically in Figure A1. The sample sizes are computed under the assumption of an infinite

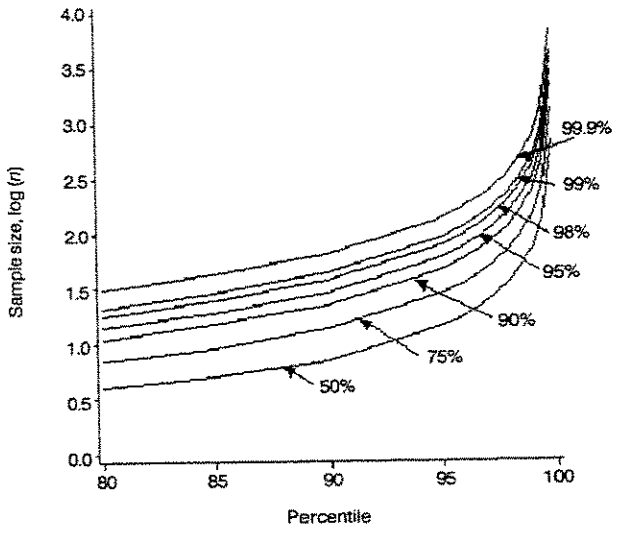

Figure A1. Required sample sizes from Table A1. The required sample size (in log 10 units) is shown as a function of the percentile to be estimated for various degrees of confidence.

Table A1. Smallest sample size required for the maximum observation to exceed 100p% of the sampled population with (1-α)% confidence

	(1−α)%						
p	50%	75%	90%	95%	98%	99%	99.9%
0.800	4	7	11	14	18	21	31
0.850	5	9	15	19	25	29	43
0.900	7	14	22	29	38	44	66
0.950	14	28	45	59	77	90	135
0.960	17	34	57	74	96	113	170
0.970	23	46	76	99	129	152	227
0.980	35	69	114	149	194	228	342
0.990	69	138	230	299	390	459	688
0.995	139	277	460	598	781	919	1379
0.999	693	1386	2302	2995	3911	4603	6905

population and are, consequently, upper bounds for sample sizes required from finite populations.

Two-sided Tolerance Intervals

Instead of covering a fixed upper (or lower) percentile of the population distribution, it may be desired to span a specified proportion of the distribution with high probability. Sample sizes based on this approach may be computed in a similar fashion, but will be larger than those given in Table A1. For further information on this approach, refer to Hahn and Meeker (1991).

Several Variables

The sample sizes computed in Table A1 ensure a prespecified probability of a fixed level of coverage of the population distribution for a single variable of interest. When several variables (e.g., different pesticides of interest) are to be covered simultaneously, it is necessary to adjust sample sizes.

To be conservative, assume that all variables are independent. When computing sample sizes for m variables, the mth root of $(1-\alpha)$ should be substituted for $(1-\alpha)$ in equation (2). For large values of m, the effect of this substitution can be quite large; for example, for $m=10$, $p=0.95$, and $\alpha=0.10$, the minimum sample size for a one-sided tolerance interval is 89, almost double the size required for a single variable.

Prediction Intervals

Another approach to sample size determination is based on the prediction of future measurements based on historical measurements. In this approach, the goal may be to predict the next observation with desired confidence or the next k future values with desired confidence. This approach requires a great deal of discussion and motivation, and consequently will not be presented here. A thorough discussion of these intervals, both parametric and nonparametric, can be found in Gibbons (1994).

Assessment and Simulations

Determining the required number of samples using the tolerance interval approach described above does not require any assumptions concerning the statistical properties of pesticide concentration distributions, and readers can make their own conclusions concerning the adequacy of the sample size recommendations based on the best professional judgment of the breakout group. But the tolerance interval approach does not account for sampling strategy (e.g., it assumes samples are taken randomly without stratification). Moreover, in practice, the nonparametric method may actually be used to estimate percentiles and their associated uncertainty. However, the determination of sample size based on the nonparametric method is not so straightforward, and definitely does depend on the statistical properties of the data under consideration. The remainder of this Appendix focuses on sample size requirements based on the nonparametric approach. Preliminary assessments are given for vulnerable surface water daily values and regional ground water sampling recommendations.

Surface Water

The data are courtesy of Dr. R. Peter Richards of the Water Quality Laboratory at Heidelberg College. The data are alachlor, atrazine, and metolachlor concentrations in surface water from 1983 through 1997 for four streams in Ohio. The stations are Rock Creek at Tiffen (89.6 km^2 drainage area), Honey Creek at Melmore (386 km^2 drainage area), the Maumee River at Waterville (16,395 km^2), and the Sandusky River near Fremont (3240 km^2 drainage area). During the peak herbicide runoff season there are nearly daily samples, sometimes even more than one sample per day. During winter, there are often no samples.

Two sets of simulation studies were conducted to evaluate the accuracy of estimated percentiles based on data obtained from several different sampling strategies. In the first study, the breakout group's best-professional-judgment recommendation of 33 samples per season was considered. For this study, data on alachlor concentrations from Honey Creek and the Maumee River for 1990 and 1995 were used. There are many missing values in these series and often some questionable values and dates. Nondetects were not indicated. To use these series, zero values were treated as zeros, missing values as missing values, and the maximum of the duplicate samples on the same day was taken as the measurement for that day.

To begin assessment of the breakout group's best-professional-judgment recommendation, the temporal profile of each series was plotted to see whether the temporal trends were roughly the same across sites and years. These profiles are shown in Figure A2. The distribution of the alachlor concentrations through time is remarkably similar for both time periods and both sites.

One way to assess the breakout group's recommendations pertaining to sample size for daily surface water values is to use these data to form some true daily series, sample from this series, obtain estimated percentiles, and compare these estimates with those calculated from the original series. To implement this idea, the 1990 Honey Creek series was used. Because there was frequently not a measurement for each day, linear interpolation between the existing values was implemented to obtain a series of 364 daily values that had the same statistical properties as the original data from the 1990 Honey Creek series. This interpolated series will serve as a reference, or "true," se-

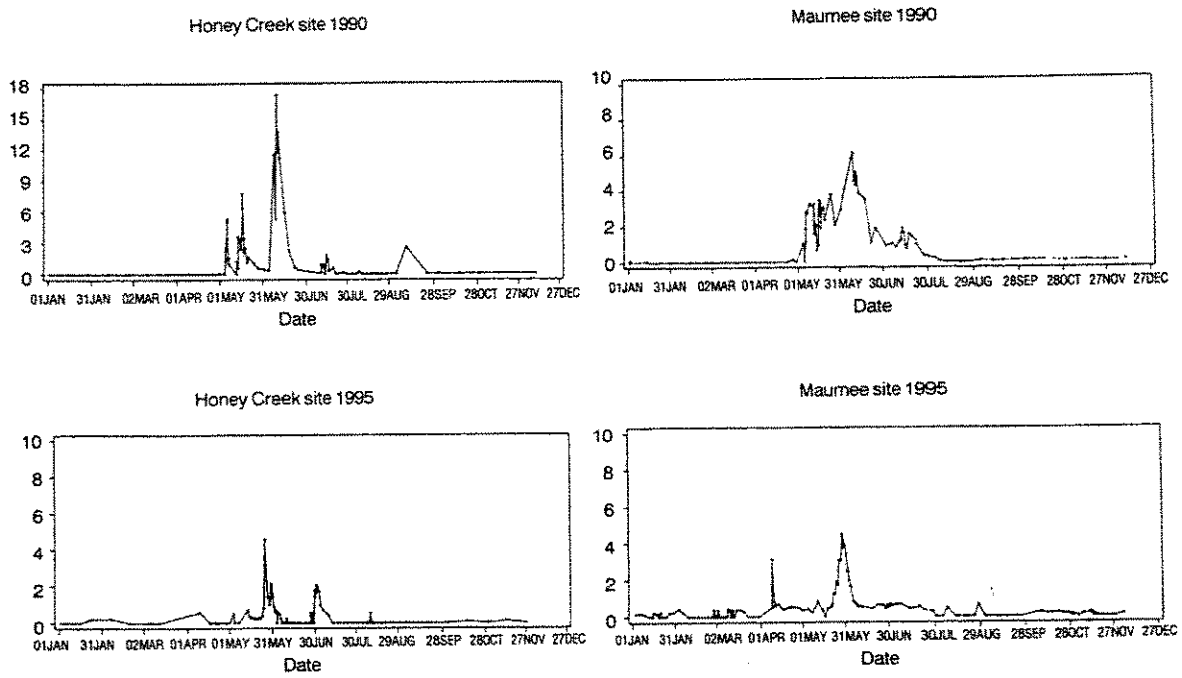

Figure A2. Temporal profiles of alachlor concentrations.

ries for the simulation study described below. Because the group did not explicitly define the seasons, after looking at the series shown in Figure A2, three seasons were defined: (1) January–April; (2) May–August; and (3) September–December.

To evaluate the sampling recommendations, it is necessary to sample 33 values from each season for each of the three seasons, estimate desired percentiles, and then compare these estimates with the "true" values obtained from the interpolated Honey Creek series. If this is done only once, however, the results will depend on the actual values obtained from just one sampling and may be biased because of sample selection. Thus, the following simulation approach was used to evaluate the sampling recommendations. First, 33 values were randomly selected from each season; then the 50th, 95th, and 99th percentiles were estimated from the sampled series; and finally, this procedure was repeated 10,000 times. This gave a distribution of values for each percentile estimate, from which the average was compared with the true values from the original series. The 5th and 95th percentiles of the simulated distribution were also calculated to form 90% "confidence intervals" in order to obtain an idea of the variability in the simulated values. The results are summarized in Table A2.

From the results in Table A2, the sample size recommendations from the best professional judgment of the breakout group seem minimally adequate, although perhaps a few more samples should be taken if estimation of the 99th percentile is of importance. Moreover, depending on the acceptable level uncertainty, more samples may be needed to reduce the uncertainty associated with the estimates of the 95th and 99th percentiles.

A second set of simulations was done to investigate the effect of different sampling strategies and to see whether the results indicated in the previous study varied with station, contaminant type, and relative amount of rainfall ("wet" years versus "dry" years). Data from the period 1992–1996 were selected from the Heidelberg College data sets described above. In this study, zero values were treated as missing observations, so some stations had periods of missing data, mostly during late fall, winter, and early spring. Because concentrations of the three herbicides are typically low during these periods of missing

Table A2. Percentiles of the distribution of alachlor concentrations (in μg/L) from the "true" values and from simulated estimates from 33 samples per each of three seasons based on the 1990 Honey Creek station alachlor concentrations

Percentile	True	Estimated	90% "Confidence Interval"
50th	0.520	0.513	(0.27, 0.81)
95th	2.666	3.594	(1.39, 5.97)
99th	11.241	11.79[a]	(5.95–17.40)

[a] The estimate may be unstable or biased, since the sample size is less than the minimum required for nonparametric estimation ($k=p(n+1)<n$, as discussed above).

values, missing values were estimated so that there were about weekly observations and no more than about 2 weeks between observations. The period of missing record was broken into four periods: September, October–December, January–March, and April. The median concentration of the available data for each period was used to estimate the missing values in each period. Piecewise linear interpolation was then used to obtain 24 values per day for each day in the year. The resulting data sets (one for each herbicide) then served as the true series for the subsequent simulation study described below.

Four sampling strategies were evaluated for each of the three herbicides at all four stations. These sampling strategies were chosen to more accurately reflect those used in practice. The first was monthly sampling. For this scenario, one observation was randomly selected from each month of the time series. For twice-monthly sampling, one value was selected randomly from the first 2 weeks of the true series, and another was selected randomly from the last two weeks of the true series. For the four-times-a-month sampling, the month was divided into four periods (days 1–7, 8–14, 15–21, and 22–end) and one observation was randomly selected from each period. For the 10-times-a-month sampling frequency, the month was divided into 10 periods (days 1–3, 4–6, 7–9, 10–12, 13–15, 16–18, 19–21, 22–24, 25–28, and 28–end) and one observation was randomly selected from each period. As before, percentiles were estimated from the sampled data and compared with those obtained from the true data sets, and then this procedure was repeated 50,000 times. Tables A3–A8 show the results for Rock Creek (representing the smallest drainage area) and the Maumee River (representing the largest drainage area). Results from Honey Creek and the Sandusky River were similar.

For instances where a yearly distribution of concentrations is of interest, Tables A5–A8 indicate that for all herbicides considered in this study, monthly or twice-monthly sampling is not sufficient for accurate estimation of higher percentiles (90th, 95th, and 99th). Four-times-a-month or weekly sampling is adequate for estimating the 50th, 90th, and 95th percentiles, but not sufficient for estimating the 99th percentile. Ten-times-a-month sampling may be needed to estimate the 99th percentile, but such estimates may not be very accurate because the standard deviations associated with the 99th percentile estimates are fairly large. Moreover, the precision associated with the percentile estimates in wet years when concentrations are higher is much lower than the precision associated with percentile estimates in dry years when concentrations tend to be lower. However, the nonparametric esti-

Table A3. Percentiles of the distributions of various herbicide concentrations (in μg/L) from the "true" values and from simulated estimates based on the values from 1992–1996 at Rock Creek

		Percentile			
Sampling Strategy	**Herbicide**	**50th**	**90th**	**95th**	**99th**
True	**Atrazine**	**0.380**	**3.065**	**5.662**	**14.179**
Monthly	Atrazine	0.380	3.441	6.857	****
		(0.012)	*(0.989)*	*(2.150)*	*(****)*
2 x Monthly	Atrazine	0.380	3.305	6.279	20.655
		(0.004)	*(0.603)*	*(1.228)*	*(11.019)*
4 x Monthly	Atrazine	0.380	3.197	5.961	17.053
		(0.000)	*(0.325)*	*(0.706)*	*(4.145)*
10 x Monthly	Atrazine	0.380	3.129	5.779	14.794
		(0.000)	*(0.148)*	*(0.319)*	*(1.866)*
True	**Metolachlor**	**0.460**	**2.136**	**3.927**	**9.489**
Monthly	Metolachlor	0.460	2.361	4.369	****
		(0.022)	*(0.671)*	*(1.485)*	*(****)*
2 x Monthly	Metolachlor	0.459	2.242	4.257	14.660
		(0.010)	*(0.389)*	*(0.845)*	*(7.724)*
4 x Monthly	Metolachlor	0.460	2.158	4.034	11.269
		(0.002)	*(0.177)*	*(0.445)*	*(2.771)*
10 x Monthly	Metolachlor	0.460	2.153	3.977	9.642
		(0.000)	*(0.073)*	*(0.225)*	*(0.801)*
True	**Alachlor**	**0.200**	**0.522**	**0.942**	**3.819**
Monthly	Alachlor	0.207	0.575	1.293	****
		(0.017)	*(0.144)*	*(0.698)*	*(****)*
2 x Monthly	Alachlor	0.205	0.544	1.099	7.042
		(0.010)	*(0.065)*	*(0.362)*	*(4.312)*
4 x Monthly	Alachlor	0.204	0.531	0.987	5.074
		(0.006)	*(0.034)*	*(0.187)*	*(1.538)*
10 x Monthly	Alachlor	0.201	0.526	0.958	4.221
		(0.002)	*(0.017)*	*(0.081)*	*(0.762)*

**** Insufficient data for estimation.

Table A4. Percentiles of the distributions of various herbicide concentrations (in μg/L) from the "true" values and from simulated estimates based on the values from 1992-1996 at the Maumee River

		Percentile			
Sampling Strategy	**Herbicide**	**50th**	**90th**	**95th**	**99th**
True	**Atrazine**	**0.344**	**3.378**	**5.686**	**10.574**
Monthly	Atrazine	0.352	3.653	6.066	****
		(0.016)	*(0.780)*	*(1.164)*	*(****)*
2 x Monthly	Atrazine	0.347	3.530	5.847	12.452
		(0.009)	*(0.427)*	*(0.699)*	*(2.559)*
4 x Monthly	Atrazine	0.344	3.523	5.780	11.662
		(0.004)	*(0.236)*	*(0.391)*	*(1.146)*
10 x Monthly	Atrazine	0.343	3.432	5.728	11.089
		(0.002)	*(0.129)*	*(0.192)*	*(0.850)*
True	**Metolachlor**	**0.417**	**2.314**	**3.791**	**7.788**
Monthly	Metolachlor	0.424	2.499	4.154	****
		(0.028)	*(0.465)*	*(0.896)*	*(****)*
2 x Monthly	Metolachlor	0.419	2.416	3.958	11.158
		(0.017)	*(0.258)*	*(0.515)*	*(4.279)*
4 x Monthly	Metolachlor	0.417	2.357	3.885	8.801
		(0.009)	*(0.134)*	*(0.266)*	*(1.304)*
10 x Monthly	Metolachlor	0.416	2.328	3.823	7.872
		(0.004)	*(0.069)*	*(0.145)*	*(0.947)*
True	**Alachlor**	**0.170**	**0.575**	**1.221**	**3.619**
Monthly	Alachlor	0.165	0.636	1.456	****
		(0.009)	*(0.170)*	*(0.570)*	*(****)*
2 x Monthly	Alachlor	0.168	0.594	1.329	5.266
		(0.004)	*(0.067)*	*(0.373)*	*(2.320)*
4 x Monthly	Alachlor	0.169	0.580	1.242	3.973
		(0.001)	*(0.032)*	*(0.213)*	*(0.525)*
10 x Monthly	Alachlor	0.170	0.576	1.216	3.700
		(0.000)	*(0.015)*	*(0.079)*	*(0.384)*

**** Insufficient data for estimation.

Table A5. Percentiles of the distributions of various herbicide concentrations (in μg/L) from the "true" values and from simulated estimates based on the values from 1994 (dry year) at Rock Creek

		Percentile			
Sampling Strategy	**Herbicide**	**50th**	**90th**	**95th**	**99th**
True	**Atrazine**	**0.435**	**2.017**	**3.347**	**5.726**
Monthly	Atrazine	0.470	3.015	****	****
		(0.092)	*(1.407)*	*(****)*	*(****)*
2 x Monthly	Atrazine	0.447	2.502	4.231	****
		(0.043)	*(0.852)*	*(1.263)*	*(****)*
4 x Monthly	Atrazine	0.436	2.114	3.749	****
		(0.019)	*(0.374)*	*(0.753)*	*(****)*
10 x Monthly	Atrazine	0.434	2.076	3.562	6.179
		(0.007)	*(0.156)*	*(0.352)*	*(0.926)*
True	**Metolachlor**	**0.501**	**1.528**	**2.257**	**4.910**
Monthly	Metolachlor	0.527	2.341	****	****
		(0.090)	*(1.229)*	*(****)*	*(****)*
2 x Monthly	Metolachlor	0.508	1.841	3.292	****
		(0.051)	*(0.525)*	*(1.327)*	*(****)*
4 x Monthly	Metolachlor	0.502	1.586	2.615	****
		(0.025)	*(0.179)*	*(0.539)*	*(****)*
10 x Monthly	Metolachlor	0.503	1.562	2.439	5.619
		(0.009)	*(0.065)*	*(0.290)*	*(1.222)*
True	**Alachlor**	**0.173**	**0.385**	**0.390**	**1.095**
Monthly	Alachlor	0.172	0.529	****	****
		(0.037)	*(0.272)*	*(****)*	*(****)*
2 x Monthly	Alachlor	0.175	0.409	0.724	****
		(0.026)	*(0.093)*	*(0.330)*	*(****)*
4 x Monthly	Alachlor	0.174	0.383	0.524	****
		(0.015)	*(0.013)*	*(0.134)*	*(****)*
10 x Monthly	Alachlor	0.175	0.386	0.390	1.316
		(0.008)	*(0.006)*	*(0.000)*	*(0.283)*

**** Insufficient data for estimation.

Table A6. Percentiles of the distributions of various herbicide concentrations (in µg/L) from the "true" values and from simulated estimates based on the values from 1996 (wet year) at Rock Creek

		Percentile			
Sampling Strategy	**Herbicide**	**50th**	**90th**	**95th**	**99th**
True	**Atrazine**	**0.382**	**4.720**	**7.203**	**19.839**
Monthly	Atrazine	0.390	9.753	****	****
		(0.052)	*(8.213)*	*(****)*	*(****)*
2 x Monthly	Atrazine	0.385	5.975	14.970	****
		(0.031)	*(1.746)*	*(10.769)*	*(****)*
4 x Monthly	Atrazine	0.380	5.044	10.352	****
		(0.016)	*(0.608)*	*(3.516)*	*(****)*
10 x Monthly	Atrazine	0.384	4.752	7.952	34.139
		(0.005)	*(0.145)*	*(1.505)*	*(16.583)*
True	**Metolachlor**	**0.477**	**3.283**	**4.913**	**12.482**
Monthly	Metolachlor	0.483	6.245	****	****
		(0.046)	*(5.284)*	*(****)*	*(****)*
2 x Monthly	Metolachlor	0.482	3.986	9.519	****
		(0.029)	*(0.967)*	*(7.190)*	*(****)*
4 x Monthly	Metolachlor	0.479	3.677	6.382	****
		(0.017)	*(0.536)*	*(1.808)*	*(****)*
10 x Monthly	Metolachlor	0.479	3.439	5.381	22.025
		(0.009)	*(0.267)*	*(0.772)*	*(11.213)*
True	**Alachlor**	**0.135**	**0.345**	**0.550**	**2.139**
Monthly	Alachlor	0.140	0.831	****	****
		(0.033)	*(0.702)*	*(****)*	*(****)*
2 x Monthly	Alachlor	0.138	0.475	1.307	****
		(0.025)	*(0.236)*	*(0.925)*	*(****)*
4 x Monthly	Alachlor	0.135	0.373	0.877	****
		(0.016)	*(0.079)*	*(0.482)*	*(****)*
10 x Monthly	Alachlor	0.135	0.362	0.632	2.962
		(0.007)	*(0.035)*	*(0.134)*	*(0.951)*

**** Insufficient data for estimation.

Table A7. Percentiles of the distributions of various herbicide concentrations (in µg/L) from the "true" values and from simulated estimates based on the values from 1994 (dry year) at the Maumee River

		Percentile			
Sampling Strategy	**Herbicide**	**50th**	**90th**	**95th**	**99th**
True	**Atrazine**	**0.417**	**1.914**	**2.452**	**3.889**
Monthly	Atrazine	0.439	2.485	****	****
		(0.076)	*(0.629)*	*(****)*	*(****)*
2 x Monthly	Atrazine	0.425	2.169	3.165	****
		(0.056)	*(0.320)*	*(0.457)*	*(****)*
4 x Monthly	Atrazine	0.417	2.016	2.957	****
		(0.026)	*(0.178)*	*(0.360)*	*(****)*
10 x Monthly	Atrazine	0.412	1.961	2.679	3.933
		(0.008)	*(0.070)*	*(0.224)*	*(0.043)*
True	**Metolachlor**	**0.441**	**1.846**	**2.333**	**3.059**
Monthly	Metolachlor	0.454	2.081	****	****
		(0.043)	*(0.511)*	*(****)*	*(****)*
2 x Monthly	Metolachlor	0.445	1.941	2.628	****
		(0.027)	*(0.379)*	*(0.318)*	*(****)*
4 x Monthly	Metolachlor	0.441	1.900	2.540	****
		(0.016)	***(0.246)***	***(0.236)***	*(****)*
10 x Monthly	Metolachlor	0.439	1.884	2.425	3.072
		(0.007)	*(0.092)*	*(0.145)*	*(0.046)*
True	**Alachlor**	**0.129**	**0.471**	**0.533**	**0.616**
Monthly	Alachlor	0.127	0.486	****	****
		(0.028)	*(0.091)*	*(****)*	*(****)*
2 x Monthly	Alachlor	0.128	0.482	0.567	****
		(0.019)	*(0.052)*	*(0.065)*	*(****)*
4 x Monthly	Alachlor	0.131	0.474	0.549	****
		(0.012)	*(0.031)*	*(0.034)*	*(****)*
10 x Monthly	Alachlor	0.131	0.473	0.540	0.673
		(0.005)	*(0.011)*	*(0.019)*	*(0.073)*

**** Insufficient data for estimation.

Table A8. Percentiles of the distributions of various herbicide concentrations (in µg/L) from the "true" values and from simulated estimates based on the values from 1996 (wet year) at the Maumee River

		Percentile			
Sampling Strategy	**Herbicide**	**50th**	**90th**	**95th**	**99th**
True	**Atrazine**	**0.355**	**4.617**	**6.480**	**10.573**
Monthly	Atrazine	0.375	6.362	****	****
		(0.066)	*(1.703)*	*(****)*	*(****)*
2 x Monthly	Atrazine	0.362	5.400	8.171	****
		(0.031)	*(0.824)*	*(1.677)*	*(****)*
4 x Monthly	Atrazine	0.356	4.971	7.151	****
		(0.014)	*(0.442)*	*(1.021)*	*(****)*
10 x Monthly	Atrazine	0.355	4.805	6.641	11.365
		(0.009)	*(0.268)*	*(0.388)*	*(1.051)*
True	**Metolachlor**	**0.451**	**3.292**	**4.510**	**6.151**
Monthly	Metolachlor	0.475	4.191	****	****
		(0.099)	*(1.161)*	*(****)*	*(****)*
2 x Monthly	Metolachlor	0.453	3.609	5.326	****
		(0.047)	*(0.676)*	*(1.198)*	*(****)*
4 x Monthly	Metolachlor	0.445	3.392	4.886	****
		(0.029)	*(0.479)*	*(0.526)*	*(****)*
10 x Monthly	Metolachlor	0.451	3.278	4.659	6.919
		(0.014)	*(0.211)*	*(0.294)*	*(1.741)*
True	**Alachlor**	**0.074**	**0.579**	**1.082**	**2.362**
Monthly	Alachlor	0.087	1.056	****	****
		(0.027)	*(0.562)*	*(****)*	*(****)*
2 x Monthly	Alachlor	0.078	0.805	1.520	****
		(0.016)	*(0.337)*	*(0.591)*	*(****)*
4 x Monthly	Alachlor	0.072	0.657	1.306	****
		(0.007)	*(0.177)*	*(0.401)*	*(****)*
10 x Monthly	Alachlor	0.073	0.597	1.147	2.501
		(0.003)	*(0.056)*	*(0.193)*	*(****)*

**** Insufficient data for estimation.

mation procedure tended to overestimate higher percentiles (sometimes substantially), particularly in smaller samples (96 samples or fewer).

Ground Water

Data used in this study are atrazine concentrations (in µg/L) obtained from sampling 770 wells in Nebraska shortly after floods in 1993 (Nebraska Department of Health and Human Services 1993). To evaluate the breakout group's sample size recommendation, two studies were conducted. First, the simulation approach as described above was used, randomly sampling 30, 50, 100, 200, 300, and 400 wells from the original 770 wells. In this study, there was no attempt to make the samples spatially representative of Nebraska. The results are summarized in Table A9.

It seems from Table A9 that the breakout group's recommendation of 200 sites per region (considering Nebraska to be a region) is adequate. Moreover, the increase in precision obtained by taking additional wells does not seem great enough to justify the additional expense, particularly for estimation of the 95th percentile. Estimation of the 99th percentile certainly requires more samples, but the gain in accuracy per 100 samples seems to be very small.

To investigate the impact of spatial scale, the original 770 wells were systematically sampled, each time increasing the spacing between the sampled wells while maintaining a reasonable coverage of the state of Nebraska. This resulted in subsample sizes of 514, 385, 193, and 97. The locations of the 770 wells and the corresponding subset of 97 wells are shown in Figure A3. The estimated percentiles and nonparametric confidence limits as obtained from equation (1) were then compared for each of these subsamples. The results are summarized in Table A10.

In Table A10, a sample size of 200 wells seems adequate for estimating the 90th percentile. However, for estimating the 95th percentile, the confidence limits from the nonparametric approach are substantially narrower for 385 wells than for 193 wells, indicating that greater precision could be obtained by taking additional wells. This is particularly true for estimating the 99th percentile, although the sample sizes required for estimating such a large percentile with high accuracy may be unreasonably large.

Spatial and Temporal Autocorrelation

Throughout this report, measurements of pesticide concentrations have been assumed to be independent. In reality, temporal or spatial autocorrelation (or both) may be present. It is not clear how such autocorrelation may af-

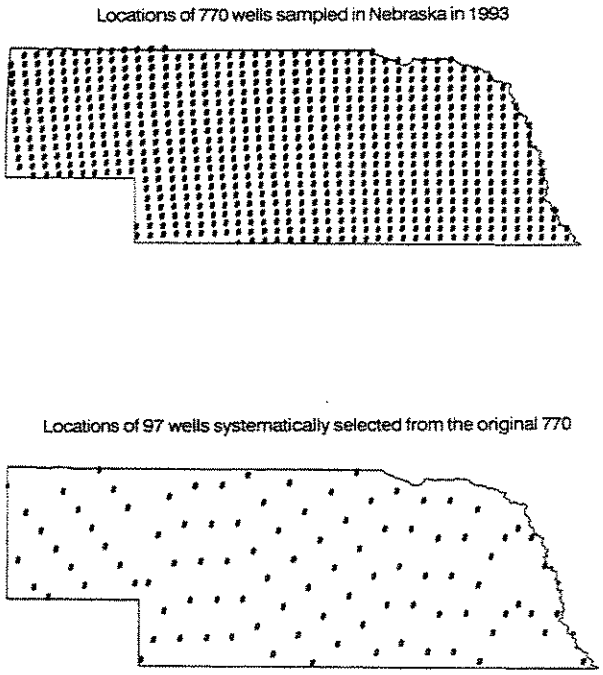

Figure A3. Locations of Nebraska ground water sampling.

Table A9. Percentiles of the distribution of atrazine concentrations (in μg/L) from the "true" values and from simulated estimates based on the Nebraska ground water data[a]

Percentile and sample size	Estimated percentile	90% "Confidence interval"
95th		
30	0.23	(0.00, 1.00)
50	0.18	(0.00, 0.74)
100	0.14	(0.00, 0.46)
200	0.12	(0.00, 0.31)
300	0.12	(0.00, 0.27)
400	0.12	(0.00, 0.25)
99th		
30	0.85	(0.00, 2.90)
50	1.19	(0.10, 4.37)
100	1.32	(0.25, 3.02)
200	1.20	(0.36, 2.53)
300	1.17	(0.49, 2.15)
400	1.15	(0.54, 2.15)

[a] The "true" values as computed from the atrazine concentrations at all 700 wells are: $\dot{X}_{0.95} = 0.13$; $\dot{X}_{0.99} = 1.05$.

Table A10. Percentiles and confidence limits for Nebraska ground water atrazine data (in μg/L).

Percentile	Confidence	Sample size	Estimated percentile	Lower confidence limit	Upper confidence limit
.90	.90	770	0.00	0.00	0.00
		514	0.00	0.00	0.00
		385	0.00	0.00	0.00
		193	0.00	0.00	0.00
		97	0.00	0.00	0.15
.90	.95	770	0.00	0.00	0.00
		514	0.00	0.00	0.00
		385	0.00	0.00	0.00
		193	0.00	0.00	0.10
		97	0.00	0.00	0.27
.95	.90	770	0.13	0.00	0.23
		514	0.11	0.00	0.27
		385	0.15	0.00	0.29
		193	0.15	0.00	0.44
		97	0.15*	0.00	2.90*
.95	.95	770	0.13	0.00	0.25
		514	0.11	0.00	0.30
		385	0.15	0.00	0.32
		193	0.15	0.00	0.72
		97	0.15*	0.00	2.90*
.99	.90	770	1.05	0.75	2.15
		514	0.80	0.74	4.37
		385	1.16	0.72	2.90*
		193	1.16	0.72	2.90*
		97	1.54*	0.44	2.90*
.99	.95	770	1.05	0.74	2.90
		514	0.80	0.72	4.37
		385	1.16	0.59	2.90*
		193	1.16	0.44	2.90*
		97	1.54*	0.44	2.90*

* The value may not be a valid estimate.

fect the uncertainty associated with the estimated percentiles (although percentile estimates should remain the same) or how this autocorrelation will affect the required sample size necessary for accurate percentile estimation. A thorough treatment of this issue in the estimation of percentiles of distributions of pesticide concentrations in drinking water is beyond the scope of this report, and may require further detailed study.

However, some guidelines can be given by borrowing on ideas from estimation of the mean (which for symmetric distributions is the 50th percentile) from correlated samples. In this case, the autocorrelation effectively serves to reduce the information content contained in the measurements, and hence larger samples are required to achieve the same degree of confidence in the estimate of the mean as with independent samples. Here, the idea of the effective number of observations may be useful. In an autocorrelated series, each observation contains part of the information contained in the previous observation. Thus, n observations in an autocorrelated series will give as much information as some lesser number, m, of observations from an independent series of measurements. (See Matalas and Langbein [1962] for theory and derivations.) The relationship between these two numbers is given by $m=nI$, where I is the information content of the sample for estimating the mean. For ease of explanation, it is assumed that the autocorrelation can be characterized by a single parameter, but multiparameter models (like those typically used to model semivariograms) could also be used. The information content is given by

$$I=[(1+\rho)/(1-\rho)-(2/n)(\rho)(1-\rho^n)/(1-\rho)^2]^{-1}. \quad (3)$$

As an example, suppose we have a sample of size 100 daily surface water alachlor concentrations that has a moderate autocorrelation of $\rho=0.3$. Equation (3) gives

$$I=[(1+0.3)/(1-0.3)-(2/100)(0.3)(1-0.3^{100})/(1-0.3)^2]^{-1}=0.5420.$$

The effective number of observations is then $m=(100)(0.5420)=54.20$, or about 54 observations. Thus, the information content in 100 autocorrelated samples with $\rho=0.3$ is the same as that contained in about 54 independent samples. Consequently, sample sizes recommended from the nonparametric method of percentile estimation or the tolerance interval approach should be increased by a factor of $1/I$. If these approaches indicate that we need about 100 samples, about 184 samples should be taken to ensure the same amount of information.

For estimating the mean from correlated samples, the effect of even moderate correlation can be severe, causing the sample size requirements to be adjusted by almost a factor of two. However, it is not clear that this effect is as strong for percentile estimates, particularly those in the tails of the distribution. As mentioned earlier, if the effect of autocorrelation on the uncertainty of the percentile estimates is believed to be an important consideration in estimating percentiles of distributions of pesticide concentrations in drinking water, then additional study must be done to fully characterize and understand any effects of autocorrelation.

Summary

This report gives a very basic overview of some of the statistical approaches to the estimation of percentiles and highlights some of the issues involved when these methods are used to estimate percentiles of distributions of pesticide concentrations in drinking water. A preliminary analysis of some "typical" data on pesticide concentrations in both surface water and ground water indicated that the sampling strategy and sample size recommendations based on the best professional judgment of the group are minimally adequate. Very accurate estimation of extreme upper percentiles (such as the 99th), however, may require a very large number of samples or a radically different sampling strategy that may not be optimal for obtaining representative distributions and estimates of lower percentiles.

Certainly the report in this Appendix is not complete, and there are several issues that may require further consideration. First, what is or what should be the goal for sample size determination? Should it be to capture a true upper-tail percentile? To span a given percentage of an underlying distribution with known confidence bounds? To estimate an entire distribution with specified confidence bounds? Or to estimate a particular percentile with a specified degree of confidence? These different methods can potentially give very different requirements for sample size. Second, what is or should be the role of randomized, stratified, probability-based sampling? Statisticians would agree that this is crucial, but implementing such a strategy can be impractical and the resulting analyses can be complicated. Third, the studies in this report were based on surface water measurements obtained from rivers in Ohio and one site of ground water measurements that was obtained during uncommon (100-year flood) conditions. It is not clear how representative these data are of pesticide concentrations in general. Finally, how will regional analyses be done? Will distributions from local estimates be combined in some way, or will locally obtained data be combined first before the distributions are estimated?

Several statistical issues have also arisen in the context of this report. First, what is known, in general, about sample size recommendations based on the goal of estimating percentiles with a known confidence using nonparametric confidence limits? Second, what is the relationship between sample size recommendations from the

tolerance-level approach and the accuracy of percentile estimates? Third, what are the statistical properties of the nonparametric percentile estimator and associated nonparametric confidence intervals? Are the coverage probabilities of these intervals accurate? How can nonparametric confidence limits be constructed for complex sampling strategies? What is the effect of autocorrelation (both spatial and temporal) on percentile estimates and confidence bounds? Finally, and perhaps most importantly, how can we use correlations between the concentrations of different pesticides to our advantage?

These issues are important ones that must be addressed as scientists struggle to better define a solid approach to risk assessments based on pesticide concentrations in drinking water. In addition to illuminating these issues and illustrating them with real data, the studies presented here also provide an overview of several approaches to sample size determination for estimating percentiles of distributions of pesticide concentrations in drinking water; provide an objective assessment of the best-professional-judgment recommendations from the group's expert panel; and give an indication of sample size requirements that will, it is hoped, be useful in future work in this fascinating but complex area of research.

References

Gibbons RD. 1994. Statistical methods for groundwater monitoring. New York: Wiley

Hahn GJ, Meeker WQ. 1991. Statistical intervals: a guide for practitioners. New York: Wiley

Matalas NC, Langbein WB. 1962. The information content of the mean. *Journal of Geophysical Research* 67:3441–3448

Nebraska Department of Health and Human Services. 1993. Domestic flood survey data, 1993. Lincoln: Nebraska DHHS

3. THE ROLE OF MODELING IN GENERATING ESTIMATES OF PESTICIDE CONCENTRATION DISTRIBUTIONS IN DRINKING WATER SOURCES

Group 2: Wendy Graham, University of Florida (Chair); Charles Crawford, U.S. Geological Survey; Conrad Heatwole, Virginia Polytechnic Institute; and Nicholas Poletika, Dow AgroSciences

3.1. Introduction

This breakout group was asked to assess what role computer modeling can play in generating estimates of pesticide concentration distributions in drinking water sources. The following questions were provided as a framework for the discussion:

- Is modeling for exposure assessment more appropriate for some assessment areas and accuracy levels than others?
- Can modeling be used to identify most vulnerable water supplies versus water supplies with average vulnerability? If so, how?
- Can modeling be used to guide monitoring? If so, how?
- What information needs to be collected to calibrate and validate existing and new models?
- What features should new or refined models incorporate?

The group discussed and deliberated each of these questions. The group agreed that models should be used in a variety of ways, including (1) generating probability distributions for exposure assessment, (2) performing vulnerability assessments, (3) designing spatial and temporal monitoring networks, and (4) extrapolating monitoring data to account for a wider range of climatic variability than observed over the data record. Summarized below are the issues and conclusions that emerged from the group's deliberations of each of the above questions.

3.2. Is modeling for exposure assessment more appropriate for some areas and accuracy levels than others?

The breakout group considered this question and discussed in detail alternative modeling scenarios that might be suitable for the three assessment areas and the three accuracy levels described in Chapter 2. Modeling approaches for surface water and ground water were discussed separately.

3.2.1. Surface Water

The breakout group first defined in more detail the three assessment scales for which probability distributions were needed: local scale, regional scale, and national scale. For modeling purposes it is most appropriate to define these assessment scales on a hydrologic basis. Once modeled, distinct hydrologic areas can be aggregated as appropriate to larger regions based on political boundaries or cropping regions.

Note that of the scales described above, local-scale modeling would most likely be a representation of a most vulnerable supply in an assessment area, if the characteristics of such a supply could reliably be identified a priori. However, as discussed in Chapter 1, defining the one most vulnerable supply in an assessment area based on site characteristics and pesticide loading and fate and transport properties is conceptually and practically very problematic, and the designation of "most vulnerable" could change as weather, pesticide use, and other dominant input factors change over the years in the assessment area. Therefore, for risk assessments that need to define the upper tail exclusively or particularly accurately, modeling of individual supplies is at times appropriate, but modeling of a *class of vulnerable supplies* might also be considered.

For purposes of the modeling discussion in this report, the group defined a highly vulnerable surface water supply to be a low-order stream in an agricultural watershed discharging into a flowing reservoir from which there was a single municipal water intake. The regional scale was defined as a higher-order river network that might have multiple withdrawal points both in streams and in reservoirs along the river network. The national surface

water supply was defined as the collection of all regional surface water supplies. There is a strong need to develop a number of representative regional scenarios from which concentration distributions for the national surface water supply can be approximated.

The group suggested that the three accuracy levels for distributions of pesticides in surface water be defined in a relative sense. Accuracy was interpreted to mean the level of confidence with which the distribution could be predicted using a particular modeling strategy. An increasing level of effort is required to move from low accuracy to high accuracy. Therefore, in order for a modeling approach to be appropriate for a particular assessment area at a particular level of accuracy, the benefit accrued from the increase in accuracy has to outweigh the cost of the increasing level of effort.

When determining model appropriateness, one must consider the current state of the art of computer technology and computer code development (i.e., available within the next 5 years) as well as the availability/measurability of input data. On this basis, the discussion of computer models was restricted primarily to physically based mechanistic models that rely more on independently measurable input parameters rather than calibration of inputs using output parameters.

Table 3.1 shows the matrix of assessment areas and accuracy levels for predicting concentration distributions in surface (or ground) water. The table indicates whether computer modeling is appropriate, marginally appropriate, or inappropriate for the given assessment area and accuracy level. The level of appropriateness is defined as follows:

Appropriate. Models are currently available that contain the important hydrologic processes for this level of accuracy, but in general have not been validated. "Off the shelf" input data required to run these models are generally available. Model validation at multiple sites for multiple chemicals is strongly recommended. After validation, accuracy of predictions from higher-category models using available data sets should improve over lower category models.

Marginally appropriate. Additional model development, calibration, and validation are required to incorporate the important hydrologic processes needed for this level. Additional field experiments and site-specific data sets are required to develop and run these models. Required model development and data gathering are recommended and should be feasible within the next 5 years.

Inappropriate. Additional model development and data gathering required to reach this level of accuracy is not likely within the next 10 years. Developing models and data sets is likely to be more costly than gathering exposure data directly from drinking water surveys.

Table 3.1 indicates that the use of computer modeling to predict concentration distributions in surface water is appropriate for the local-scale/vulnerable assessment areas at the low- and medium-accuracy levels and marginally appropriate if high accuracy is needed. The use of computer modeling is marginally appropriate (given the current state of the art) for the regional and national scales at low- and medium-accuracy levels and inappropriate at these scales if high accuracy is needed in the distributions. Summarized below are some of the more important issues for each assessment area and accuracy level.

3.2.1.1. Local Scale/Vulnerable–Low Accuracy

The local-scale scenario for a surface water supply was conceptualized as a highly vulnerable watershed having spatially uniform soil and chemical application practices and a 100% cropped area. For a low level of accuracy and a high level of conservatism, this scenario could be simulated by incorporating minor modifications into any number of existing field-scale or, preferably, watershed models. Several groups have prepared reports and informal documents that summarize existing watershed model capabilities, for example, the U.S. Environmental Protection Agency (EPA) Office of Water (1995), the National Agricultural Chemicals Association (1994), the EPA Office of Pesticide Programs (an example of an informal background document can be found at http://www.epa.gov/pesticides/SAP/1998/July/matrix.htm, January 1999), and the FIFRA Exposure Model Validation Task Force (an example of an informal report can be found at http://www.femvtf.com/recommen.htm, January 1999). Many of the models described in these reports are capable of generating distributions over time of the pesticide concentrations in surface water. Using climatic variability as the major driver of uncertainty, one can generate probabilistic concentration distributions.

This scenario would provide a conservative estimate, since most people do not drink water from a 100% cropped,

Table 3.1. Recommendations for the use of computer models for predicting pesticide concentration distributions in ground water

Accuracy level	National	Regional	Local
High	Inappropriate	Inappropriate	Inappropriate
Medium	Marginal	Marginal	Appropriate
Low	Appropriate	Appropriate	Appropriate

100% treated watershed with soils vulnerable to generating surface runoff. However, for many small rural watersheds, the conservatism may be warranted if documented cases can be found where 100% cropped area is approached. In cases where it is not warranted, it would be relatively easy to incorporate a representative crop area factor, with percent crop treated to add realism. Because the scenario is soil specific, crop specific, and pesticide management specific, and the receiving reservoir is flowing rather than a static pond, it should provide some improvement over GENEEC (GENeric Estimation of Environmental Concentrations model) predictions.

3.2.1.2. Local Scale/Vulnerable–Medium Accuracy

To advance to a medium level of accuracy for a highly vulnerable surface water supply, coupling the watershed model with a Geographic Information System (GIS) is necessary to allow for deterministic spatial distributions in land use and soils and for routing of runoff, sediment, and pesticides from the edge of fields through the watershed to the reservoir. For a given land use, the breakout group felt that uniform pesticide use and crop management practices could be assumed at this level of accuracy. Modeling a spatially distributed watershed would allow dilution with water running off uncropped areas and varying travel times from different pesticide sources. In general, in-stream transformation and degradation processes could be neglected for small watersheds at this level of accuracy. Runoff rates and pesticide concentrations entering the reservoir would be variable over time because of variable climate and pesticide application patterns. The reservoir could be assumed to be well mixed and flowing, and dissipation and degradation of the pesticide in the reservoir could be modeled. Thus, using weather patterns as the major driver of uncertainty, one could generate probabilistic concentration distributions in the reservoir.

Existing watershed models (e.g., see documents cited in the Local Scale/Vulnerable–Low Accuracy section above) could be relatively easily adapted for this purpose. Accuracy in these model predictions, however, would be limited by the ability to define accurately the land uses, soil types, pesticide use, and cropping practices, as well as all the required processes and input parameters. Furthermore, all of these models require additional validation to predict the fate and transport of pesticides.

3.2.1.3. Local Scale/Vulnerable–High Accuracy

To move to a higher level of accuracy for the local scale, highly vulnerable scenario probabilistic distributions of the spatial processes must be accounted for. For example, crop rotations could be taken from a random distribution representative of the area, soil properties might be spatially variable within a field, and chemical use and application dates might also be spatially variable. At this accuracy level, more sophisticated reservoir models that accounted for water demands, storage outflow relationships, evaporation, and precipitation processes could be included. Landscape factors that could account for buffer zones and pesticide losses between the edge of a field and a stream could also be incorporated. Again, using spatially uniform (because of a small watershed area) but temporally variable weather patterns as the major driver of uncertainty, one could generate probabilistic concentration distributions in the reservoir.

The watershed models summarized in the documents cited in Section 3.2.1.1 could be relatively easily adapted for this purpose. However, as previously indicated, accuracy in these model predictions would be limited by the ability to define accurately all of the required processes and input parameters, and confirmation of accuracy requires further model validation.

3.2.1.4. Regional–Low Accuracy

The regional scale was defined as a higher-order river network that might have multiple withdrawal points both in streams and in reservoirs along the river network. There is a strong need to develop a number of representative regional scenarios for which computer modeling might be attempted. Thus, a particular regional system would be modeled after an actual higher-order river network with defined surface water withdrawal points. At the low level of accuracy, land use, soils, and chemical use would be spatially variable but deterministic. The river network and reservoirs could be modeled with rudimentary storage, routing, and in-stream/in-reservoir processes. Because of the increased size of the river network, weather should be modeled as both spatially and temporally variable. The distribution of pesticide concentrations in the surface drinking water would be compiled over both space and time. In an exposure assessment, the pesticide concentration of the water withdrawn from surface water supplies in the region could be weighted by the population served by that supply.

Significant additional model development and data gathering are required to obtain regional concentration distributions even at the low-accuracy level. However, several of the currently available river network models described in the documents cited in Section 3.2.1.1 could serve as possible starting points. The major problem at the regional scale is the need for extensive data to validate and, possibly, calibrate existing models, but these data are generally not available. Therefore, computer modeling is only marginally appropriate for this category at present, but may become feasible in the near (about 5 years) future.

3.2.1.5. Regional–Medium Accuracy

To move to a medium level of accuracy for the regional scenario, there is a need to develop new, more physically based models that do not rely so heavily on calibration of inputs using output data. At the medium level of accuracy, accounting for probabilistic distributions of the spatial processes may be important. For example, as discussed above, crop rotations could be taken from a random distribution representative of the area, soil properties might be spatially variable within a field, and chemical use and application dates might also be spatially variable. More sophisticated reservoir models that accounted for water demands, storage outflow relationships, evaporation, and precipitation processes could be included. Landscape factors that could account for buffer zones and pesticide losses between the edge of the field and the stream could also be incorporated. Spatially and temporally variable weather patterns, spatial distributions of soils, cropping patterns, and surface water supply locations would generate a distribution of pesticide concentrations in the regional surface drinking water supply.

Models are not generally available that can adequately simulate the processes described above. Thus, the use of modeling for this category is inappropriate at the current time, but may become marginally appropriate within the next 5 to 10 years. Computer models will be inappropriate to generate high-accuracy regional concentration distributions within the foreseeable future.

3.2.1.6. National–All Accuracy Levels

The national surface water supply was defined as the collection of all regional surface water supplies. If computer models could be used to generate distributions at any accuracy level for all regional surface supplies (or some representative subset of surface water supplies), the results could theoretically be aggregated up to obtain the same-accuracy national concentration distribution. At the present time, this is marginally appropriate at a low-accuracy level and increasingly inappropriate at higher-accuracy levels. At high-accuracy levels, monitoring pesticide concentration distributions may be more cost effective than modeling.

3.2.1.7. Surface Water Summary

In summary, surface water modeling is most appropriate for local-scale, highly vulnerable small agricultural watersheds, and becomes less appropriate at larger regional scales, primarily because of the difficulty of accurately defining the required model processes and parameters. If the highly vulnerable modeling scenarios described above were conducted at progressively higher accuracy levels, compounds that failed the early-tier screening might pass under the higher-tier screens.

Significant effort needs to be devoted to collecting data to validate the models selected for use, even at the local-scale, lowest-accuracy level. The EPA is considering funding a project to conduct a preliminary model evaluation and comparison effort for existing watershed models (e.g., SWAT [Soil and Water Assessment Tool], RIVWQ [River Water Quality model], BASINS [Better Assessment Science Integrating Point and Nonpoint Sources model]) with existing data gathered from watersheds on the order of 100 square miles in size (see http://www.epa.gov/pesticides/SAP/1998/July/final1.pdf, January 1999). This is a good first step.

An effort also should be made to define regional surface water supplies for which surface water models should be developed, verified, and calibrated. Specifically, the group recommended that a pilot study be initiated to demonstrate the regional-scale modeling strategy. The purpose of this pilot study would be to develop, validate, and calibrate a regional model at multiple sites to evaluate the scalability of the available models to large (e.g., 100 to 10,000 square mile) high-order watersheds. This pilot study should develop methodologies to use these regional models to extrapolate concentration distributions to other unmodeled, unmeasured sites so that national distributions can be constructed.

Significant effort needs to be devoted to data collection specifically targeted for model validation at all scales. Thus, well-designed monitoring studies should be initiated both to characterize the input parameters that are needed for the models and to measure surface water pesticide concentrations in order to quantify the models' prediction accuracy. Note that the type of monitoring required to parameterize and validate predictive models is not necessarily the same as the type of monitoring required to conduct risk assessments. Risk assessment only requires estimating exposure to pesticides through drinking water; however, predictive modeling requires understanding all of the processes controlling concentration of pesticides in drinking water. It is not clear that both objectives can be efficiently handled within a single monitoring study.

An empirical regression modeling approach similar to the SCI-GROW (Screening Concentration In GROund Water model) ground water model could hold promise as a conservative early-tier (i.e., low-accuracy) screen for surface water. However, these types of models would likely not be appropriate if medium or high accuracy is required, as when estimating complete distributions of pesticide concentrations. Effort and resources would probably be better spent to develop and parameterize the more physically based process models described above.

3.2.2. Ground Water

The breakout group again proceeded by first defining in

more detail the three scales of assessment for which it is desirable to generate probability distributions: local scale, regional scale, and national scale. As noted previously, local-scale modeling would most likely be a representation of a highly vulnerable supply in an assessment area. However, as discussed in Chapter 1, defining the one most vulnerable supply in a use region is problematic, and the designation of "most vulnerable" could change as dominant input factors change over the years in the assessment area. Therefore, for risk assessments that need to define the upper tail exclusively or particularly accurately, modeling of individual supplies is at times appropriate, but modeling of a *class of vulnerable supplies* might also be considered.

For purposes of the modeling discussion in this report, the local-scale scenario of a highly vulnerable ground water supply was defined as a single homeowner's domestic water supply well in a shallow surficial aquifer underlying highly permeable soils in a rural agricultural area. The regional ground water supply was defined as the collection of all individual and municipal water supply wells withdrawing from a particular (geologically distinct) surficial or confined aquifer system. The national ground water supply was defined as the collection of all regional ground water supplies. The levels of accuracy for which distributions were needed, and levels of appropriateness for model use, were interpreted in the same manner as for surface water.

Table 3.2 shows the matrix of assessment areas and accuracy levels for predicting concentration distributions in ground water. Indicated within the table is an assessment of whether computer modeling is appropriate, marginally appropriate, or inappropriate for the given assessment area and accuracy level.

Table 3.2 indicates that the use of computer modeling to predict concentration distributions in ground water is appropriate for all assessment areas if distributions are needed at low-accuracy levels. However, the use of computer modeling is inappropriate (given the current state of the art) for any assessment area if high accuracy is needed in the distributions. For medium levels of accuracy, the use of computer models is most appropriate for local ground water supplies and least appropriate for defining the distributions for national ground water supplies. Summarized below are some of the more important issues for each assessment area and accuracy level.

3.2.2.1. Local Scale–Low Accuracy

The local, vulnerable ground water supply was defined as a single homeowner's domestic water supply well in a shallow surficial aquifer underlying highly permeable soils in a rural agricultural area. For a low level of accuracy and a high level of conservatism, a particular vulnerable soil, crop, and pesticide management scenario could be simulated using any one of a number of existing vadose-zone models. Reports that summarize existing vadose-zone model capabilities have been prepared by a number of groups, including the FIFRA Exposure Modeling Work Group (see, e.g., National Agricultural Chemicals Association 1994), the FIFRA Scientific Advisory Panel to the EPA Office of Pesticide Programs, and the FIFRA Exposure Model Validation Task Force (see the websites given in Section 3.2.1.1). Many of the vadose-zone models described in these reports are capable of generating distributions over time of the concentration of vadose-zone water leaching into the saturated zone. Using climatic variability as the major driver of uncertainty, one can generate probabilistic concentration distributions.

This scenario provides a conservative concentration estimate, since people are not drinking leachate water directly. However, for a situation in which the entire well capture zone lies within a surficial aquifer receiving recharge from a 100% cropped area, the conservatism may be warranted. Because the scenario is site specific for weather, soil characteristics, crop management practices, and pesticide use, it should provide some improvement from the generic SCI-GROW prediction that is based on limited data.

3.2.2.2. Regional–Low Accuracy

The regional ground water supply was defined as the collection of all individual and municipal water supply wells withdrawing from a particular surficial or confined aquifer system. For the regional ground water supply, low-accuracy conservative distributions can be predicted by embedding the vadose-zone model described above into a GIS system. In this modeling scenario, the recharge area for an entire aquifer system would be mapped in the GIS system. Each unique soil-crop-pesticide management scenario could be modeled independently using the vadose-zone model, allowing for climatic variability. A concentration distribution for the region would then be compiled over both space and time from the predictions of the con-

Table 3.2. Recommendations for the use of computer models for predicting pesticide concentration distributions in surface water

Accuracy level	National	Regional	Local
High	Inappropriate	Inappropriate	Marginal
Medium	Marginal	Marginal	Appropriate
Low	Marginal	Marginal	Appropriate

centration in vadose-zone water leaching into the ground water. This scenario provides less conservative predictions than the local-scale low-accuracy scenario by allowing for consideration of less vulnerable soils and noncropped regions within the aquifer recharge area. At this level of accuracy, well capture zones are assumed to be influenced solely by adjacent treated fields (i.e., regional-scale lateral flow of pesticide residues in ground water is neglected).

3.2.2.3. National–Low Accuracy

The national ground water supply was defined as the summation of all regional ground water supplies. Thus, if low-accuracy assessments can be conducted for all regional ground water supplies (or some representative subset of ground water supplies), the results could theoretically be aggregated up to obtain a low-accuracy national concentration distribution.

3.2.2.4. Local Scale/Vulnerable–Medium Accuracy

To advance to a medium level of accuracy for the local-scale vulnerable ground water supply, some kind of saturated zone flow and transport model is required for the individual well capture zone. Modeling the entire ground water capture zone for a vulnerable well allows consideration of transformation and degradation in the ground water, dilution with water recharged from uncropped areas in the capture zone, and varying travel times from different pesticide sources. The group felt that the temporal concentration distributions from the low-accuracy vadose-zone model described above could be used as the upper boundary condition for a saturated-flow pesticide-transport model. Leaching rates and pesticide concentrations would be variable over time because of variable climate and pesticide application patterns. Thus, using the upper-boundary condition as the major driver of uncertainty, one could generate probabilistic concentration distributions in a vulnerable well.

Existing saturated-zone ground water flow and transport models could be relatively easily adapted for this purpose, for example, MODFLOW/ MT3D. (A description of this and other U.S. Geological Survey [USGS] ground water modeling software can be found at http://water.usgs.gov/software/, January 1999.) Accuracy in these model predictions, however, would be limited by the ability to define accurately the well capture zone and all the required processes and input parameters.

3.2.2.5. Regional–Medium Accuracy

Conservative distributions can be predicted by using the GIS-based vadose-zone model described above for the entire aquifer recharge area to define the time distribution of pesticides entering the upper boundary of a regional ground water flow and transport model. A concentration distribution for the region can then be compiled over both space and time. Such a distribution could be used in an exposure assessment by weighting the pesticide concentration of the water withdrawn from each individual and municipal water supply well in the region by the population served by that well. Accuracy in these model predictions would be limited by the ability to define accurately in space and time the aquifer recharge area and all the required processes, boundary conditions, and input parameters.

3.2.2.6. National–Medium Accuracy

If medium-accuracy assessments could be conducted for all regional ground water supplies (or some representative subset of ground water supplies), the results could theoretically be aggregated up to obtain a medium-accuracy national concentration distribution. However, the practicality of accurately modeling even a representative subset of the national ground water supplies is questionable. Thus, in the near future, computer modeling is marginally appropriate for this category.

3.2.2.7. All Assessment Areas–High Accuracy

Moving to a high level of accuracy using computer models to generate the pesticide concentration distributions in ground water supplies might be theoretically feasible using a densely discretized, fully transient, three-dimensional model coupled to a vadose-saturated-zone regional flow and transport model. However, obtaining accurate parameters for such a model is currently infeasible, and the computer costs of running such a model would be prohibitive. Thus, computer modeling is probably inappropriate for this category, and it would be less expensive and more accurate to measure pesticide concentrations in a large sample of drinking water taps to obtain the desired distribution. Obviously this is feasible only for estimating distributions for existing pesticides.

3.2.2.8. Ground Water Summary

In summary, modeling is most appropriate for local-scale ground water supplies where low accuracy in the distribution is acceptable. Modeling becomes less appropriate as more accuracy is desired over larger regions. The primary limitation at the medium level of accuracy is gathering the site-specific data needed to parameterize the models, whereas at the high level of accuracy, both model development and data gathering are significant limitations. Generally, if the low-accuracy and medium accuracy modeling scenarios described above were progressively run for compounds that failed the SCI-GROW screening, the

higher-tier screens might be required less frequently.

As with the surface water models, significant effort needs to be devoted to collecting data to validate the models selected for use even at the local, vulnerable, lowest-accuracy level. The bounds of application and the expected accuracy of model predictions need to be more clearly defined before pesticide concentration distributions can be estimated with any confidence. An example of this type of model validation exercise is the work of the FIFRA Exposure Model Validation Task Force on the validation of PRZM (Pesticide Root Zone Model) version 3.12 using data from prospective ground water studies (see http://www.femvtf.com, January 1999). It should be noted that this task force initially also intended to evaluate the GLEAMS (Ground Water Loading Effects of Agricultural Management Systems) model. However GLEAMS has been dropped from consideration because it is no longer supported by the U.S. Department of Agriculture (USDA). It is recommended that some entity (EPA or others) should assume responsibility for continued support and development of models that are accepted for regulatory and/or exposure assessment use in order to allow for their continued utility.

The key to using models to estimate medium-accuracy estimates of concentration distributions at the regional and national levels is to select representative regional scenarios to model. Thus, an immediate effort should be made to define a suite of representative regional (aquifer/crop/ pesticide/weather) scenarios for which the medium-accuracy ground water models described above could be developed. The MUSCRAT standard scenarios (for a description of MUSCRAT, see http://www.femvtf.com/working.htm, January 1999) should be considered a potential starting point for the development of representative regional scenarios. A significant amount of new data must be collected on pesticide use, land use, soils, geology, and chemical properties for the representative regional scenarios selected for model application. Also, some additional model development will be required. Thus, an important first step is to develop a means of supporting model development and implementation.

3.3. Can modeling be used to identify highly vulnerable water supplies versus water supplies with average vulnerability? If so, how?

Modeling can and should be used to identify relative vulnerability rankings for ground water and surface water supplies. The low-accuracy modeling scenarios described above, or even simpler leaching and runoff index models, are already being used for this purpose. An example of this type of work is found in a report by Kellogg et al. entitled *Twenty-Five Year Trend in the Potential for Environmental Risk from Pesticide Leaching and Runoff from Farm Fields* from the USDA. This report is published at http://www.nhq.nrcs.usda.gov/land/pubs/pesttrend.html/ , January 1999. Also available on this website are maps of the Pesticide Leaching Index and the Pesticide Runoff Index for the continental United States that were created using the Soil-Pesticide Interaction Screening Procedure (SPISP). Maps of the potential for exceeding drinking water standards at the edge of the field and the bottom of the root zone are based on the National Pesticide Loss Database [AU: Should this be included in the References?].

Studies like the example given above represent appropriate uses of currently available models for vulnerability assessment and for selecting sites for further monitoring and/or modeling. Use of models for these purposes acknowledges the implicit assumption that the models are more reliable for relative comparison than for the actual numbers they produce. There is some concern that there generally has been little verification of the accuracy of these vulnerability assessments using field-scale measurements. Thus, evaluating the accuracy of the vulnerability assessment maps using more detailed models is a consideration, especially in regions where field data are available. However, it is difficult to justify a recommendation that more effort and resources be invested in verifying the accuracy of vulnerability maps. In general, resources would be better spent in developing and validating more detailed process models.

3.4. Can modeling be used to guide monitoring. If so, how?

Modeling can and should be used to guide ground water and surface water monitoring. For example, with the use of the vulnerability maps discussed above, surface and ground water monitoring sites can be distributed between more vulnerable and less vulnerable areas as appropriate, based on pesticide use region. Furthermore, more detailed temporal surface and ground water modeling can help to discern when to initiate sampling, how often to sample, and how long to sample a particular well, stream, or reservoir after a specific pesticide application or cropping season.

3.5. What information needs to be collected to calibrate and validate existing and new models?

It must be emphasized that monitoring data needed for model development and calibration are significantly different from data needed to do an exposure assessment. Furthermore, the data gathering for model development is put at risk if it is "piggy-backed" on a monitoring study designed for exposure assessment. For example, informa-

tion on ground water levels, surface water stage and discharge, and suspended sediment concentration is necessary for process modeling, but not for exposure assessment. Existing drinking water wells can be used for exposure assessment, but are generally not suitable for gathering data needed for modeling purposes. Similarly, event-based surface water monitoring is critical for model development and calibration, but not necessarily for chronic exposure assessment. It must also be emphasized that data requirements also differ based on the goals and objectives of the modeling study. For example, different types of data may be needed for the calibration of an accepted established model for use with a particular chemical of interest in a particular geographic location than for the development and validation of new or refined models. Furthermore, the spatial resolution of ancillary data that will be required depends on the model and the accuracy level. In short, a rational and specific plan must be developed for all monitoring studies that have specific objectives based on the proposed use of the data. A pilot study should be initiated to develop a monitoring strategy that is specifically designed to obtain the data sets that would be needed to develop, validate, and calibrate the physically based models described in this report.

Other models requiring further development are physically based models to extrapolate measured data to unmeasured sites and to extrapolate monitoring data over time to account for a wider range of climatic variability than observed over the data record. However, to conduct theses types of extrapolations, significant additional ancillary data are required along with pesticide concentration data (e.g., weather, land use, cropping patterns, pesticide use, soil characteristics, slopes, hydrogeology). In general, efforts to extrapolate pesticide concentration data over either space or time will fail unless the monitoring scheme was designed for that purpose and the required ancillary data were collected a priori.

Although simulations using independently measurable input parameters are preferred, there may be instances where calibration of certain processes can improve accuracy in predicting pesticide concentrations. For example, HSPF (Hydrologic Simulation Program–FORTRAN) usually requires adjusting stream hydrology and sediment transport input parameters to obtain the best possible agreement with water flow and sediment measurement data. This allows simulation of pesticide fate and transport in the basin in a noncalibrated mode with respect to the chemical, combined with the most realistic hydrology and sediment transport estimates from the model. Assuming that acceptable accuracy in the pesticide concentration predictions for one chemical is obtained, the calibrated system can then be used to predict concentrations of other chemicals in the same watershed, given identical weather and similar pesticide use.

3.6. Which features should new or refined models incorporate?

The above sections discuss the various processes that should be incorporated into physical models to estimate pesticide concentration distributions for exposure assessments at the various accuracy levels. Summarized below are the major features that should be incorporated into new or refined models to meet these objectives:

- incorporation of mitigation practices such as buffer strips, detention ponds, or grassed waterways,
- incorporation of the effects of tile drains and other drainage devices,
- incorporation of preferential flow processes in vadose-zone models,
- incorporation of improved mechanisms to route edge-of-field water, pesticides, and sediments to streams,
- incorporation of more sophisticated in-stream and in-reservoir processes,
- development of a probabilistic framework to systematically account for climatic and parameter spatiotemporal variability and uncertainty, and
- development of effective ways to represent the degree of spatial variability required in model parameters to obtain a good representation of the pesticide concentration distribution at the various levels of accuracy.

References

National Agricultural Chemicals Association. 1994. Report on primary, secondary, and screening models for pesticide registration. Washington, DC: NACA

US Environmental Protection Agency, Office of Water. 1995. Compendium tools for watershed assessment and TMDL development. Washington, DC: EPA

4. DATA AGGREGATION ISSUES

Group 3: Douglas Crawford-Brown, University of North Carolina (Chair); Timothy Barry, U.S. Environmental Protection Agency; Catherine Eiden, U.S. Environmental Protection Agency; Raymond Layton, DuPont Agricultural Products; and David Miller, U.S. Environmental Protection Agency

4.1. Introduction

Traditionally, the U.S. Environmental Protection Agency (EPA) has evaluated the safety of pesticides based on a single-chemical, single-exposure-pathway scenario. When risk assessments were performed in this manner, additional or secondary exposure pathways were not considered. Risk assessments that incorporated only exposures through food as the primary pathway of exposure might miss significant exposures through such secondary routes as dermal absorption or drinking water. In addition, exposures were treated as independent events; i.e., one individual was assumed to be exposed to one pesticide through one pathway at a single point in time. In the real world, exposures to pesticides do not occur as single unconnected events, but rather as a series of sequential or simultaneous (and potentially dependent) events through a variety of pathways that are linked in time and place.

In 1996, Congress passed the Food Quality Protection Act (FQPA), which amended both the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) and the Federal Food, Drug, and Cosmetic Act (FFDCA). FQPA defined a new standard for pesticide safety as "reasonable certainty that no harm will result from *aggregate* exposure to the chemical residue, including all anticipated dietary exposures and all other exposures for which there is reliable information" (italics added). These laws mandated that the U.S. Environmental Protection Agency (EPA) register pesticides and set tolerances based on a safety determination and that it consider such *aggregate* exposures in its decision-making process.

Aggregate exposure assessment has been defined by the International Life Sciences Institute (ILSI) as "a process for developing an estimate of the extent of exposure of a defined population to a given chemical by all relevant routes and from all relevant sources" (ILSI 1998). Aggregate exposure assessments consider all routes of exposure (e.g., through air, water, soil, and food) and all uses (e.g., agricultural uses and home and garden uses) and ideally should incorporate relevant information regarding the co-occurrence of events and exposures. An illustrative diagram of relevant considerations and pathways in aggregate exposure assessment is shown in Figure 4.1.

In considering aggregate exposure from pesticides, one must consider contributions from drinking water (*dw*), food (*f*), soil (*s*), inhalation of air (*a*), and dermal absorption as shown in Figure 4.1. The calculation of risk usually begins with the calculation of the exposure, which is the average rate at which a person takes the pesticide into the body (in units such as mg/day) divided by the body weight. The general equation for exposure is:

$$\text{Exposure (mg/kg/day)} = (C_{dw}IR_{dw} + C_fIR_f + C_sIR_s + C_aIR_a + C_sIR_s)/BW, \quad (1)$$

where C_i is the concentration of the pesticide in medium I (e.g., in units of mg/L), IR_i is the intake rate (ingestion or inhalation) or contact rate for dermal absorption for medium i (e.g., in units of L/day or cm^2/day), and BW is the body weight of the exposed individual (e.g., in units of kg). For drinking water, the subscript i is *dw*, and the contribution in equation (1) is given by the first term on the right-hand side. (Note that this simplified equation is provided to introduce the concept of exposure assessment and does not reflect how current aggregate exposure models calculate food or residential exposures; for further discussion, see ILSI [1998.]) In the following discussion, it is assumed that modeling and/or monitoring data for drinking water are available to perform an aggregate exposure assessment.

Given these data, the charge to the breakout group was: *How should modeling and monitoring data for drinking water be incorporated into an aggregate exposure assessment?* In other words, given a scientifically reasonable procedure to provide estimates of potential concentrations of pesticides in drinking water, how should these estimates be incorporated into an aggregate assessment?

Any proposed methodology for incorporating drinking water (or any media) into an aggregate exposure assessment must recognize that each contributing pathway in equation (1) is characterized by spatial and temporal

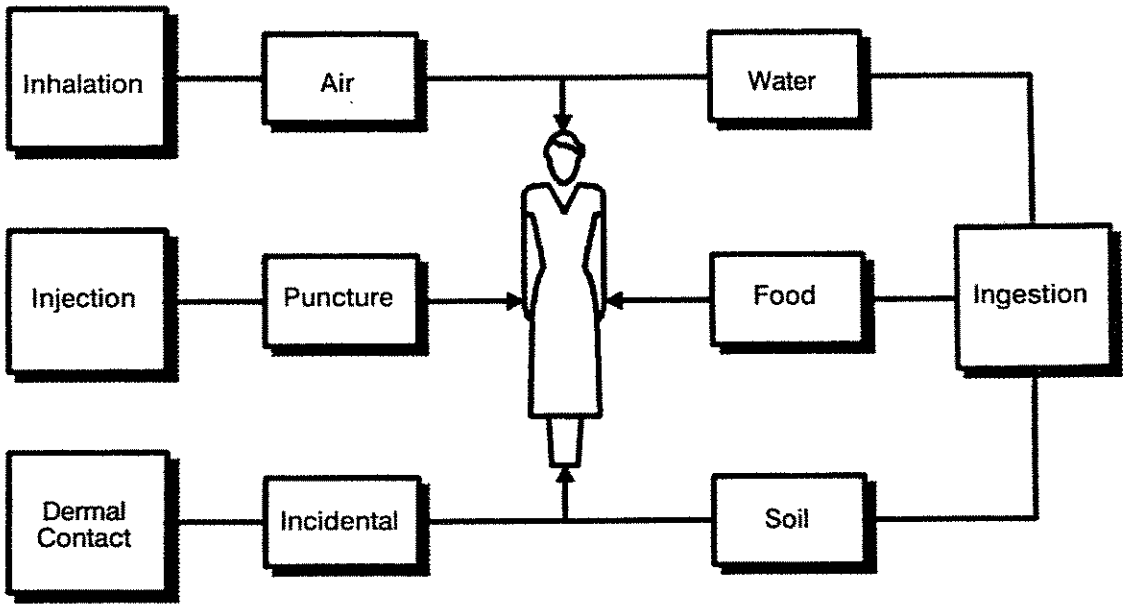

Figure 4.1. Routes of exposure and intake to be considered in a complete aggregate exposure and risk assessment. In practice, one or more routes (e.g., incidental exposures and injections) may be ignored if they are insignificant.

variability in the concentrations within the relevant environmental media. In other words, the concentration varies across a locale and varies with time within a locale. These separate contributions to variability must be appropriately combined to yield the composite variability of aggregate exposure; a conceptual diagram showing this process is displayed in Figure 4.2. For the pathway of drinking water, there will be variability in:

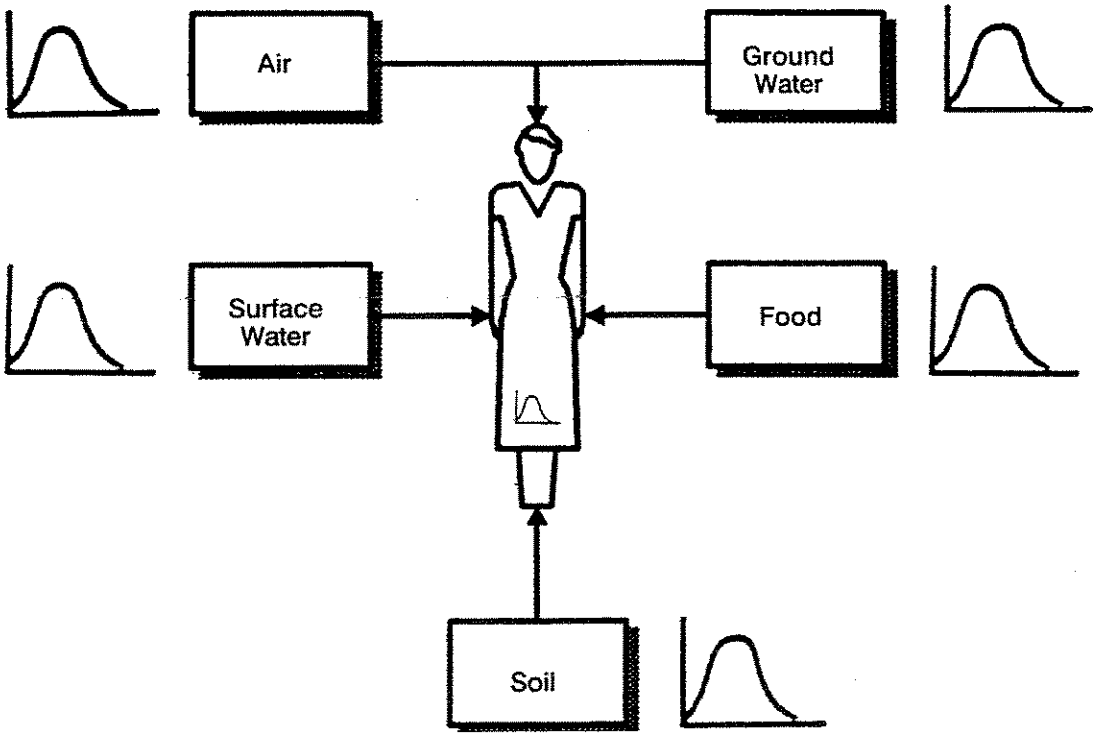

Figure 4.2. An example of aggregate exposure assessment to be performed for pesticides. Probability density functions are defined for each of the exposure pathways, including variability of concentration in the medium and variability of exposure factors. These are combined through Monte Carlo analysis to yield the composite variability distribution for aggregate exposure. All instances of correlation must be included.

- the concentration of the pesticide in the water across different water sources at a given time (spatial variability),
- the concentration of the pesticide in the same source water of the same water source supply across time (temporal variability),
- the treatment system across systems and time, resulting in variability in the concentration of the pesticide in the finished water at different times,
- the distribution and mixing that takes place between the finished water and the tap, resulting in variability of the pesticide in the tap water among different water customers using the same supply at a given time, and
- the exposure factors for individuals consuming water (e.g., there is intersubject variability in the ingestion rate of water per unit body mass and temporal variability in these same exposure factors throughout the year [Finley et al. 1994]).

In addition to these general considerations concerning the temporospatial variability of the concentration of pesticides in drinking water and variability in exposure factors, several other specific considerations place requirements on the information generated by the drinking water exposure assessment:

- It may be necessary to develop exposure information on *different subpopulations*. For example, concern over sensitive subpopulations or minority groups may require that the information be generated for these subpopulations separately from the information generated for the general U.S. population.
- It may be necessary to develop exposure information for *different time intervals*. For example, concern for acute effects may require information on daily exposures, whereas concern for chronic effects may require information on annual or lifetime exposures.
- It may be necessary to develop exposure information for *different seasons*. For example, the distributions may change significantly between winter and late spring or between a period when pesticides are used and one in which pesticides are not used (the focus usually is on use season rather than calendar season). This consideration will be most important for acute effects where even brief periods of high exposure may be unacceptable.
- It may be necessary to develop exposure information for *different treatment groupings*. This might occur, for example, because data or model predictions are available only for source water but not for finished water. In such a case, it will be necessary to adjust the source water concentrations by a "treatment, mixing, and distribution factor" to account for the effect of treatment, mixing, and distribution on water concentrations at the point of exposure (e.g., the tap). These factors, in turn, are likely to be a function of source water characteristics such as ground water versus surface water, large supply versus small supply, and rural versus urban.
- It may be necessary to develop information on the variability of concentrations in water bodies, in source water (i.e., at the point of intake into the treatment and distribution system), and/or at the tap. The reason for focusing on these different points of water use is that risk management strategies could focus on any of these points, for example, on watershed protection for the first, on moving the point of intake for the second, or on treatment in the third.

The need to provide these separate bodies of information on variability of exposure within the population across space and time will depend on risk management goals (e.g., the goal to consider environmental equity) and/or on the calculations that must be performed following generation of the distributions of exposure in the population (e.g., the need to correct for the effect of water treatment, mixing, and distribution in different source waters). The general rule is that producing separate distributions (showing variability of pesticide concentration in drinking water for all of the above considerations) is complicated and should occur only if a policy goal or subsequent calculation requires it. This is why it is necessary to understand the uses to which information on variability of exposure will be put in risk management decisions when deciding on the strategy for generating those distributions.

The evaluation of exposure through drinking water across time, geography, and population subgroup is important, but it also is important to remember that exposure to drinking water is only one part of the overall aggregate exposure. Exposure to pesticides in drinking water must be evaluated in context with other types of exposure (e.g., food, residential). For example, it does not make sense to combine an exposure distribution for drinking water in Iowa with the exposure distribution determined for fire ant control in and around homes in Texas. The possible coincidence of each of the exposure routes must be taken into account when doing an aggregate risk assessment.

4.1.1. Tiered Approach to Aggregate Exposure Assessment

It was agreed that any attempt to incorporate drinking water exposures into an aggregate risk assessment should incorporate a series of tiers, or screening procedures, to assess whether drinking water exposures could potentially contribute in any substantial way to aggregate risk. A tiered approach will produce the simplest and most efficient sampling and modeling requirements. Progression through the tiers is expected to result in more accurate and realistic estimates of concentrations in drinking water. Two approaches were proposed, each of which matches the level of effort required with the necessary (and required) degree of confidence in the estimate.

The first approach assumes that the risk manager simply wants to be sure that the secondary pathways and other sources of exposure will not cause the aggregate exposure to exceed allowable levels under any reasonably anticipated circumstances. Such a decision may be considered an issue of screening, since the goal is to identify pesticides for which there is no reason to suspect that exposure to all pathways and all sources might produce an unacceptable risk. If there does appear to be the possibility of an unacceptable aggregate risk, a more detailed assessment of aggregate exposure may be needed. The goal is reached by using estimates of exposure from secondary pathways and sources that are so conservative that there is virtually no possibility that they will be exceeded by anyone in the population. If the total exposure from both the primary pathway and these conservative estimates of the secondary pathway is below a level of concern, there is no need to perform a detailed assessment of the secondary pathways.

In this approach, the resulting drinking water exposure provided to the risk manager *does not* represent an accurate estimate of aggregate exposures for most of the population, although it might be representative for some maximally exposed individual. Still, the risk manager can be highly confident that the resulting aggregate assessment will yield protective management options (which does not preclude the possibility that the policy is overprotective). This approach requires that the drinking water exposure assessment address two questions. First, for what concentration of a pesticide in drinking water can we be very certain that there is no one in the exposed population receiving drinking water at a higher concentration, assuming that this concentration could be found somewhere at some time? And second, is that a concentration that causes concern in light of aggregate exposure? The last part of the first question is provided to ensure that physically unrealistic values are not selected. It is important to remember that any drinking water exposure estimate developed under the first approach represents a hypothetical extreme and should be characterized as such in any exposure assessment.

A second possible approach assumes that the risk manager wants to more fully characterize the distribution of secondary exposures across the exposed population or subpopulation of interest, and to combine these with the primary exposures to produce a distribution for aggregate exposures in the population. The reason for doing this might be to obtain an accurate estimate of population risk that considers the probability distribution associated with each exposure pathway and combines them in a mathematically appropriate manner. In contrast to the first approach, in this second approach an accurate representation of the distribution of aggregate exposure in the exposed population can be produced, assuming suitable data and/or models are available. This approach provides the most information, but it also is the most information intensive because it is necessary to generate a distribution of exposures for the primary routes and/or sources, generate a distribution for the secondary routes and/or sources, determine any correlation between these two distributions, and produce the distribution of aggregate exposures in the population.

In the following discussion, it is assumed that estimates of exposure may be produced either through monitoring data, through models, or through some combination of the two. The ideal source of information clearly is monitoring data, assuming those data are reliable and representative. As indicated in the modeling section of this report (Chapter 3), there are some cases where modeling will be useful or even necessary. This might be so because monitoring data are nonexistent or, for some reason, are considered to be unreliable for determining the exposure profile. For example, no monitoring data would be available for a newly developed product. Another example might be when raw water-monitoring data are available for a compound, but modeling may be needed to estimate the effect of simple filtration or chlorination on the exposure profile for a pesticide.

4.2. General Considerations for Aggregate Assessment

The basic concept underlying all aggregate exposure assessments is that exposure occurs to a random individual. The integrity of the data concerning this exposed individual must be consistently maintained throughout the aggregate exposure assessment. Each of the individual "subassessments" must be linked back to the same person, and the aggregate intake must reflect dietary, drinking water, and residential intakes that are for the *same individual* at the *same time*, in the *same place*, and under the *same demographic conditions*. In other words, the aggregation must be *simultaneously* temporally, spatially, and demographically specific, i.e., the data must agree in time, place, and demographic characteristics. It would be incorrect, for example, to simulate an individual's exposure by randomly selecting a dietary contribution from an entire population's distribution of dietary exposures, combining that distribution with a randomly selected drinking water contribution from an entire distribution of drinking water exposures, and combining these two independently selected contributions with a third randomly selected residential contribution from an entire population's distribution of residential exposures (ILSI 1998).

One difficulty in performing aggregate exposure analyses is that the exposure pathways may or may not have independent distributions. For example, the region with the highest drinking water exposure might be the north-

eastern United States because of orchard use. The region with the highest residential exposure might be the southeastern United States because of household use. Performing an assessment by just simply combining the two distributions will lead to erroneous conclusions. Thus, it is important that an aggregate exposure assessment estimate the exposure that a randomly selected individual will receive from all relevant routes and pathways under a given set of internally consistent circumstances. Such an estimate of individual exposure and risks may vary in space and time (e.g., in different places and on different days). An aggregate exposure assessment will seek to portray the range of individual exposures that may be received in a well-defined population of such individuals as a distribution of exposures, reflecting the influence of varying individual characteristics (e.g., age, sex, ethnicity, place of residence, occupation).

Developing realistic aggregate exposure and risk assessments requires that the appropriate temporal, spatial, and demographic exposure factors be correctly assigned and consistently maintained. Specific considerations should include:

- time (duration, frequency, and seasonality of exposure; seasonally based pesticide residues in food; frequency of residential pest control reflecting housing location and type),
- place (location and type of home; urbanization; watershed or aquifer characteristics; region; regionally specific drinking water concentrations of the pesticide being considered), and
- demographics (age; sex; sex- and age-specific body weights; reproductive status; ethnicity; personal preferences, behaviors, and characteristics).

Aggregate exposure and risk assessments are first completed for individuals, who are then combined to develop distributions of exposure to subpopulations and populations.

When considering exposure events, it is important to keep in mind the concept of conditionality. Conditionality means an exposure event on one day may also produce or have some effect on exposure events on subsequent days. For example, if a pesticide application and a rainfall event coincide to produce a flush of pesticide residues into a local drinking water supply source by surface water, the individuals served by that supply may receive exposures on subsequent days as the pesticide is distributed in the supply if the pesticide is not removed by treatment. As the pesticide concentration declines with time, subsequent exposures (on subsequent days) would decline as well. Another example might be if a private ground water well is contaminated with a persistent pesticide on any given day; household members would then be exposed to similar concentrations in the drinking water for an extended period of time thereafter. In both of these examples of exposure events, short-term exposures may be assessed based on single-exposure events or estimations of 1-day concentrations of the pesticide in the supply; however, some sort of time-weighted average of the pesticide concentration is needed to accurately assess long-term exposures to the chemical. This concept is discussed in more detail in sections 4.4.10 and 4.4.11.

In addition, the assessment must appropriately incorporate linkages or correlations/associations (which can be either positive or negative) between use events. For example, in some cases the use of one product may affect the likelihood of using another product. This might be true with respect to products used for flea control: an indoor fogger, lawn care product, and a flea product for a pet might be more likely to be used simultaneously. In other cases, the products may serve essentially the same purpose, such that the use of one will almost certainly preclude the use of the other. As another example, places of residence should be linked or otherwise correlated to water source. It is much more likely, for example, that a residence located in a rural site in the Midwest will have a private well as a source of the household water supply than a residence in an urban location in the Northeast. In this case, the location of the residence must be linked with the source of the water supply to appropriately incorporate real-world situations and ensure that unrealistic or unlikely combinations are appropriately discounted.

4.3. A Practical Framework for Analysis

The focus of this section is on the techniques of data aggregation, particularly how existing or anticipated information on concentrations of pesticides in drinking water may be used to estimate the contribution of direct water ingestion of pesticides to aggregate exposure to pesticides. Consideration of exposure through drinking water requires the measurement or prediction of the distribution of pesticide concentrations in finished drinking water (e.g., the distribution showing the fraction of the population or subpopulation using drinking water at any given concentration of the pesticide). Specifically, this section provides a discussion of how existing or anticipated monitoring data and/or modeling approaches could be incorporated into an aggregate exposure assessment for pesticides in drinking water. It describes potential conceptual approaches and methodologies for incorporating data on drinking water concentrations of pesticides into an aggregate assessment as well as issues associated with this incorporation.

In practice, any approach for assessing drinking water contamination must consider the localized or regional nature of the water supply and the fact that the levels of contamination (if any) will be driven by factors such as

use practices and cropping patterns specific to a given pesticide as well as various factors such as climate, hydrology, and soils. Large, randomized national surveys for pesticide contamination as conducted by the EPA (the National Well Survey [EPA 1990]) provide information on drinking water quality on a national basis. However, these types of surveys are by design not intended to provide information specific to a given pesticide's use pattern and potential "hot spots" of contamination in local water supplies. The latter type of information comes from a stratified, random monitoring design, which incorporates some bias toward sampling in high-use areas surrounding vulnerable water supplies.

It is commonly believed that for the majority of pesticides, contamination of drinking water supplies is expected to be localized as opposed to widespread. This poses a difficult problem in designing drinking water monitoring surveys: we need to recognize that some individuals may be highly exposed whereas the majority are not, and design monitoring programs to provide information on that potentially exposed subpopulation as well as the remainder of the population. To dismiss either group would be inappropriate, and to impose the high exposures of the few on the many would be false.

The breakout group discussed a practical, general framework or approach to incorporating existing and anticipated data on pesticides in drinking water into aggregate exposure assessments. Table 4.1 describes the types of data that could be used to generate exposure assessments for the three assessment areas (national, regional, and site specific [most vulnerable]) and the confidence levels associated with each type of assessment. The group agreed that to conduct probabilistic exposure assessments, a full distribution of pesticide concentrations in drinking water based on random (or stratified random) sampling ideally is needed. In lieu of a full distribution of pesticide concentrations in drinking water, a realistic approximation of that distribution could be used in a distributional analysis. The use of data from targeted, focused monitoring in which samples are collected only from "highly vulnerable" sites cannot in a truly statistical sense be used in probabilistic exposure assessments. Exposures based on the results of a subsample of pesticide concentrations in drinking water from a targeted (nonrandom)

Table 4.1. Geographic scales of assessment and levels of certainty for estimating the contribution of pesticides in drinking water to aggregate exposure

	Assessment area of interest		
Certainty	**National**	**Regional**	**Site-specific (most vulnerable)**
High	Anticipated stratified, random sampling of finished drinking water. Sites would be categorized as to vulnerability, and the sampling design would oversample randomly from the more vulnerable sites. Results can be used for probabilistic exposure assessment.	Anticipated stratified, random sampling of finished drinking water targeted to a specific region. Sites would be categorized as to vulnerability, and the sampling design would oversample randomly from the more vulnerable sites. Results can be used for probabilistic exposure assessment.	Focused, targeted monitoring in vulnerable finished drinking water for specific compounds. Results can be considered to be representative of worst-case exposures at other vulnerable sites, but cannot be extrapolated regionally or nationally. Could be used for site-specific distributional analysis of exposure.
Medium	Existing monitoring data from various sources distinguished as representing either drinking or source water data using regional models[a] to interpolate in areas where no monitoring data exist. Source water[b] results would need to be modified for the effects of treatment where warranted.	Existing monitoring data from various sources distinguished as representing either drinking or source water data using regional models[a] to interpolate in areas where no monitoring data exist. Source water[b] results would need to be modified for the effects of treatment where warranted. Extrapolation of results to other regions is not advised.	Existing monitoring data from various sources distinguished as representing either drinking or source water data using site-specific models[a] to interpolate between sites. Source water[b] results would need to be modified for the effects of treatment where warranted.
Low	Assessment not advisable.	Combination of models and monitoring data on source water and ground water, i.e., the NAWQA data set, generic water quality data sets.	Screening-level assessment using model estimates compared against a standard or risk-based limit for a pesticide in drinking water in light of other aggregate exposures.

[a] Models must be validated against existing monitoring data, and not used in isolation of monitoring data. Uncertainty in model predictions must be factored into exposure assessments.
[b] Source water is water that could potentially be used as a source of drinking water.

survey from a most vulnerable site would have to be carefully characterized as to what they really represent (i.e., high-end exposures). As such, data from the most vulnerable sites could only be extrapolated to other similar most vulnerable sites. The group agreed that it would be inappropriate to extrapolate exposures based on site-specific (most vulnerable) data to regional or national levels. As such, the breakout group developed recommendations for conducting aggregate assessments at three different levels:

- site-specific distribution of concentrations (typically the most vulnerable water supply or collection of most vulnerable supplies),
- regional or watershed distribution of concentrations, and
- national distribution of concentrations.

These distributions of concentrations in drinking water would typically be given as a probability density function (PDF) or cumulative distribution function (CDF) showing the fraction of individuals, utilizing a small subset of the most vulnerable sites, exposed to a concentration of any particular value (see Figure 4.3 for an example). Following the two decision approaches outlined in Section 4.1.1, one could specify (1) a concentration of the pesticide in a water supply that is likely to represent a reasonable worst-case scenario of exposure in the distribution across this subset and (2) the complete variability distribution of pesticide concentrations in drinking water for the subset. As also discussed in the previous section, these two bodies of information might be needed for any of several disaggregated subsets such as special subpopulations, use season, daily/annual averages, and/or treatment groupings.

For each of these assessment areas, it is desirable to define data and analytical needs for a relatively high degree, a moderate degree, and a relatively low degree of certainty in the distributions. It is important to remember that a low degree of certainty in drinking water concentration estimates may simply mean that the gross methodologies used to develop that estimate provide admittedly extreme overestimates of exposure and thus do not directly imply that we have a low degree of confidence that health effects are not occurring.

Levels of confidence associated with each level of assessment would vary depending on the types of data available and used in the assessments. The levels of confidence associated with the different geographic scales of assessment using existing and anticipated data sets and model predictions are given in broad, descriptive terms as high, medium, and low.

Based on this information, the group developed a matrix (Table 4.1) that describes the types of analyses or sampling that would be desirable to generate probabilistic distributions for the three assessment areas and certainty levels. Table 4.1 can be viewed as a tiered process used to distinguish between pesticides that can be screened out as a drinking water concern at a minimal level of analysis and pesticides that may require a full exposure and risk analysis. For those pesticides that fail the screening tiers and require a detailed risk assessment, the preferred approach to perform the dietary (food and water) portion of an aggregate exposure assessment is to combine a probabilistic drinking water exposure assessment with a probabilistic food exposure assessment performed by a Monte Carlo analysis. Progression from the lower (less refined) to the higher (more refined) tiers is expected to produce less conservative and more accurate/realistic estimates of drinking water concentrations to which individuals are actually exposed. Each of the tiers is described in additional detail below.

Low-confidence assessments at all geographic levels. Existing screening models can appropriately be used as screens to identify pesticides for which there is no drinking water concern. If it can be demonstrated that the highly conservative pesticide concentration estimates provided by the model never approach a drinking water concentration level that would lead to concern, one can be reasonably certain that drinking water is not likely to be a significant source of exposure. Although the certainty in these model-generated surface and ground water concentration estimates from existing models are admittedly very low, there is extremely high confidence that these concentrations are rarely, if ever, exceeded in drinking water. Therefore, although limited in use, existing surface and ground water models with some modifications toward improvement as discussed in Chapter 3 can be used strictly for screening purposes only (see Section 4.4.4 for further discussion on the use of models).

The group agreed that validation of existing models against existing monitoring data and subsequent refinement of the models is warranted to improve the use of the models as screening tools and before their results could be used directly in a quantitative exposure assessment to set policy or regulate. As discussed in Chapter 3, on the role of models in estimating pesticide concentrations in drinking water sources, once existing models are validated and improvements are made to improve accuracy, models could be used for site-specific assessments where low accuracy is required. Regional ground water assessments may include the use of model results along with available ground water monitoring data. Regional surface water assessments using available surface water data sets (i.e., the National Water Quality Assessment program [NAWQA]) or other nonrandom, nontargeted surveys or compilations of water quality data could be used where low certainty is acceptable. The use of surface water models beyond the site-specific/local scale was not advised. The group agreed that national assessments of drinking

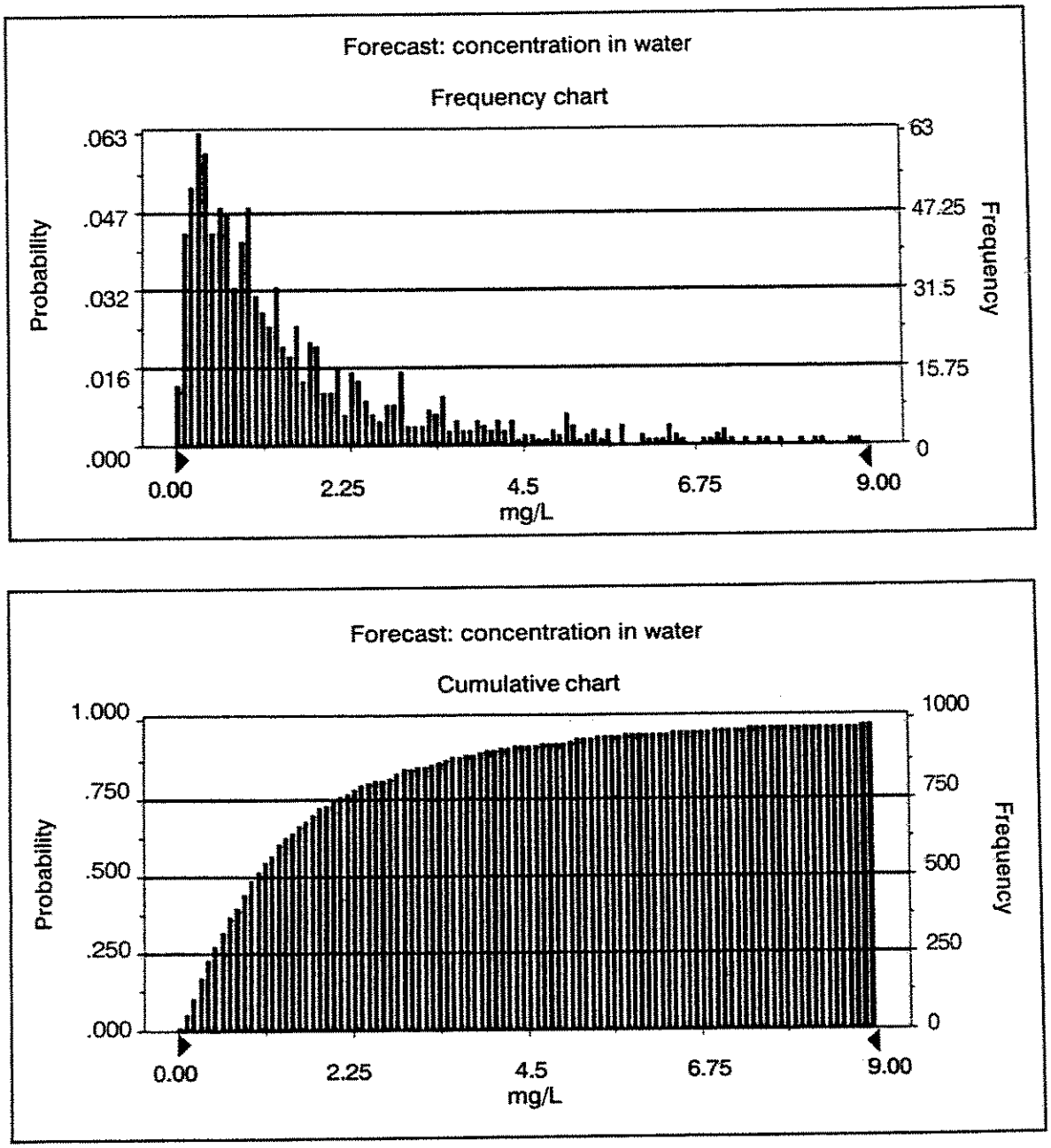

Figure 4.3. An example of probability density function (top figure) and associated cumulative probability function (bottom figure) that must be generated for a probabilistic aggregate exposure assessment. One of each must be generated for each exposure pathway, and any correlation between two or more probability density functions must be established to ensure a proper estimate of the variance of the aggregate exposure.

water exposure based on site-specific or regional assessments of low accuracy and confidence would require extrapolating "at your own peril," so to speak, and was not advised.

Medium-confidence assessments at all geographic levels. Where it cannot clearly be demonstrated that a pesticide represents a de minimis source of exposure compared with food and residential exposures, it is necessary to proceed to a more refined level of analysis requiring proportionally more resources and time to develop. The group agreed that where medium confidence is required for an aggregate exposure assessment, some sort of ad hoc meta-analysis that combines monitoring data for both drinking water and source water with the use of models to interpolate between sites would be the most likely approach. Any models used would have to comply with the validation and modification steps discussed in the modeling chapter of this document before they could be used

with medium confidence. Little or no additional or "targeted" collection of data is seen as necessary at this level of analysis which would provide "moderate" certainty of no significant exposures through drinking water. This approach may instead necessitate the use of "opportunity data," such as water quality monitoring data from state departments of health, prospective ground or surface water studies, retrospective studies by the U.S. Geological Survey (USGS), EPA, and various states and independent parties, including Safe Drinking Water Act (SDWA) data and EPA's Storage and Retrieval (STORET) data, or other heretofore untapped resources. However, the monitoring data used would have to be designated as finished drinking water versus source water, and any source water monitoring data used in lieu of finished drinking water data would need to be modified for the effects of treatment where warranted. In the absence of this modification, we would assume that source water residues carry through to the tap unchanged. Any uncertainty associated with the use of these data would have to be described. The group agreed that this approach could be used at all geographic levels of assessment—site specific, regional, and national—with medium confidence. It is believed that data from these sources could potentially be combined and compared to a calculated standard risk-based limit for a pesticide in drinking water in light of aggregate exposures.

High-confidence assessments at all geographic levels. To achieve a high level of accuracy and confidence in an exposure assessment for pesticides in drinking water, it is anticipated that additional drinking water monitoring data will be needed. This level of analysis requires considerably more resources, effort, time, and expense than those required in the two previous levels and is the first tier at which probabilistic analyses would ordinarily be conducted. The analyses at this level would provide statistically valid drinking water data at a regional level (making sure that sufficient emphasis is placed on obtaining samples from vulnerable sites), and would generally be performed only when a "low certainty" assessment and/or "medium certainty" assessment was unable to demonstrate that drinking water concentrations were not of concern in the aggregate risk assessment.

To obtain this type of data, statistically robust, stratified, random sampling was advised. The sampling design would need to accommodate the hypothesis that for pesticides there will be hot spots of contamination based on individual pesticides' use patterns and characteristics of individual water supplies. The group agreed that for exposure assessments at the regional (and national) levels of assessment where a high level of confidence is required, pesticide concentrations in drinking water at the tap or the point of distribution after treatment were warranted. Source water could be used for exposure assessments only where a high level of confidence is required, if the effects of treatment across the country for a wide variety of treatment systems are known. To simplify the assessment, tap water or the closest thing to it, finished drinking water at the point of distribution, is preferred. All agreed that the current approaches using models and targeted surveys to assess drinking water exposure to pesticides are not statistically robust for the majority of pesticides, and the results of such approaches could not be used in probabilistic exposure assessments. The group also agreed that some a priori knowledge regarding the variability within the distribution of pesticide concentrations in finished drinking water would be needed to design even a simple monitoring plan. In general, the group agreed that stratified, random samples of concentrations in drinking water are not available for most pesticides at this time, but a long-term goal should be to obtain this type of data.

4.3.1. Specific Considerations for Vulnerable Site Assessments

For site-specific assessments at sites considered to be most vulnerable where a high level of confidence is required and no statistical inference is necessary, focused, targeted monitoring in vulnerable drinking water supplies could be used. Analyses for a "vulnerable site" generally would be expected to be conducted at the county or small watershed level such that the drinking water data collected would represent a homogeneous population that could therefore be pooled and sampled in a probabilistic manner. Sampling should occur over a time period sufficient to ensure a representative sample and, although conducted year-round, should nevertheless be concentrated during those seasons in which high pesticide concentrations in drinking water might be expected (e.g., stratified random sampling). Naturally, the probabilistic assessment would weight the samples such that the frequencies were represented in a valid manner.

If such an analysis revealed that exposures were not of concern, one could confidently conclude that drinking water concentrations represented a minimal source of exposure for both the local (vulnerable) population and certainly for the U.S. population as a whole. We note that if exposures were higher than desired, a risk manager could require that mitigation actions be undertaken for the affected locales; potentially, a higher (regional) analysis may need to be performed for which additional (more extensive) statistically designed sampling in a larger geographic region would be required.

4.3.2. Specific Considerations for Regional Site Assessments

If a previous "vulnerable" site probabilistic analysis revealed that individuals in vulnerable sites were potentially exposed to high levels of pesticides in their drinking water, a regional-type analysis might be recommended.

This analysis would be expected to consist of a wider area of interest (generally at the multiple-state level) and involve a much more detailed and involved stratified sampling plan. Sampling could be stratified on a number of variables expected to be important in determining whether a sampling location is contaminated (e.g., size of community served, rural versus urban/suburban, ground versus surface versus mixed water source, shallow versus deep well, extensive versus minimal treatment, private versus community source). Naturally, seasonality would also be reflected, and because an aggregate assessment is being performed, the size of the population would also be incorporated.

The breakout group agreed that for exposure assessments at the regional (and national) levels of assessment where a high level of confidence is required, measurements of pesticide concentrations in drinking water at the tap or the point of distribution after treatment were warranted. Source water could be used only for exposure assessments where a high level of confidence is required, if the effects of treatment across the country for a wide variety of treatment systems are known. To simplify the assessment, tap water or the closest thing to it, finished drinking water at the point of distribution, is preferred.

4.3.3. Uncertainty Characterization in Exposure Assessments

For any exposure assessment at any of the above levels of assessment and their associated levels of confidence, characterization of the assessment as to uncertainties in that exposure assessment is integral to the assessment and must be included. Because perfect information on temporal and spatial variability within the distribution of pesticide concentrations in drinking water is not available, some resolution will be lost in any exposure assessments and will have to be captured through an uncertainty analysis (Crawford-Brown 1997; Morgan and Henrion 1990). Exposure assessments must reflect the current state of knowledge. In addition, there needs to be a clear differentiation between exposure estimates produced on a "best guess" or "not likely to exceed" basis and exposures based on values in which we have confidence. The group agreed there was a need to make a comparison between model results and monitoring results to develop an "uncertainty factor" that may be applied to drinking water exposures based on model estimates.

Most exposure assessments revolve around the concept of average exposures (based on arithmetic means) and upper-bound exposures for random individuals. However, exposure assessments need to be realistic in the sense that there should be actual individuals potentially exposed at the levels considered. The group agreed that there needs to be some idea of a national distribution of pesticide concentrations in drinking water in order to place exposure estimates based on modeling or monitoring of most vulnerable sites in perspective with respect to the full distribution of exposures.

Site-specific assessments. Because of the "hot spot" issue relative to drinking water contamination by pesticides (as discussed above), site-specific assessments may often be targeted to drinking water supplies that are considered more vulnerable to pesticide contamination. Exposure assessments based on data from a subset of highly vulnerable drinking water systems singled out for analysis cannot be extrapolated to other systems. No statistical inference can be made about exposure beyond the specific sites sampled; the only information available will be the concentration for a pesticide within a given system, exposures related to that system, and individuals served by that system.

Characterization of exposures based on this type of worst-case data is especially critical. Results from any site-specific, targeted monitoring efforts in most-vulnerable areas would have to be characterized as nonrandom but representative of high-end exposures. In a targeted monitoring plan, the samples are not random and no clear conclusion about probability of occurrence or exposure beyond the specific area targeted for sampling can be made from them. However, the group discussed that as a theoretical construct, in the case where there is no exposure of concern at the most vulnerable sites, the assumption could be made that no exposures of concern are expected at the less vulnerable sites. This approach would easily be open to criticism.

Regional assessments. When data from individual drinking water supplies are combined for regional assessments, the degree of randomness in the samples must be considered. If the samples were not chosen randomly, but are from targeted surveys, expert judgment will be necessary to characterize the regional exposure assessment. For example, if only data from worst-case sites are combined within a region for a regional assessment, the assessment does not represent drinking water exposure for the whole region; on the contrary, it represents drinking water exposure for the most highly exposed individuals (portion) of the region. As another example, if data from existing monitoring databases are used, those data will have to be described as representing source water, drinking water, or neither (shallow ground water, streams, or ditches), and any correlation between the areas/sites sampled and pesticide use in those areas will have to capture the uncertainty associated with the data.

National assessments. The same concepts described above apply to capturing uncertainty and characterizing a national assessment for drinking water exposures. The group agreed that any national assessment would have to be characterized as to issues relating to localized areas of concern (i.e., hot spots).

4.3.4. Data Availability

For pesticides, contamination of drinking water is highly localized and regionalized, driven by pesticide use, cropping patterns, climate (e.g., rainfall), and landscape (e.g., slope). A large, randomized national survey was conducted (EPA 1990) that provides some information on drinking water quality on a national basis. However, the EPA (1990) survey is restricted to ground water supplies and may not provide an accurate representation of the upper tails of the distribution needed in developing aggregate exposure for purposes of risk assessment. More complete data sets are available for regional and small-scale water sources, such as the data set represented by the USGS's NAWQA program, but these data focus on source water rather than finished or tap water. Interpreting the findings of the USGS study is not straightforward; for example, because the study did not consider routes of exposure other than water, not all of the pesticides and metabolites detected have standards under SDWA, and the USGS sites were not necessarily drinking water sources or even potential drinking water sources.

Currently, regulatory-driven monitoring studies focusing on most vulnerable settings do not provide randomized samples across such sites and are specific to source water rather than finished or tap water. At best, these might be used to supply upper-bound estimates on exposure through source water, acting purely as a screening methodology rather than providing an accurate representation of the variability distribution needed for a complete aggregate exposure assessment.

For a few pesticides some data are collected under the SDWA on a national basis, and this effort will be expanded in the future through EPA's Unregulated Contaminant Monitoring Rule. However, rural drinking water wells are excluded (a serious omission for the purposes of developing aggregate exposure information in agricultural regions).

Potential data that should be explored for combination into a single database can be categorized as follows:

Prospective field-scale studies. Small-scale prospective ground water studies are required to be performed (on an "as needed" basis) before registration or reregistration of pesticides. Data generated from these studies include shallow ground water and soil cores from treated fields and soil water collected from the unsaturated zone by lysimeters. For surface water, data could be collected on small streams and rivers adjacent to treated fields. Either set of data might provide additional information on the upper tails of the source water distribution, although such data will not represent realistic exposures in drinking water, since these data typically do not focus on drinking water concentrations (they should be used only if they can be related through modeling to surface water concentrations at actual points of exposure). The use of this type of data is not encouraged for use in drinking water exposure assessments without proper characterization.

Retrospective studies. There are a number of past and ongoing water quality monitoring efforts conducted by the USGS, EPA, and various states and institutions. These are not designed to track a particular pesticide from point of application, but to provide information on general pesticide occurrence (which is the issue of interest here). Examples are the NAWQA database, drinking water samples collected under the SDWA (with 25 pesticides, although only community water supplies are considered), and studies conducted using STORET data.

Pesticide-specific data. Some data sets supply detailed information on a single pesticide. The Acetochlor Registration Partnership (ARP), the Atrazine Survey (examining finished water for both surface and ground water supplies), and the state Pesticide Management Plans (PMPs, for ground water supplies) are candidates. Although these studies were not designed specifically to provide a national or regional random, stratified sample, the data will still be useful in providing an indication of possible drinking water exposure.

4.4. Challenges and State of the Art

Sections 4.1 through 4.3 consider theoretical aspects and practical implementation of an aggregate exposure assessment that would include exposure to pesticides in drinking water. This section reviews the limitations of the existing data and models and the potential effects of those limitations on the task of incorporating drinking water exposures into aggregate risk assessment in more detail. The discussion is constructed around a series of questions most relevant to assessing the strengths and limitations of the ideal approach.

4.4.1. What are the methods available to integrate analyses of drinking water data with other exposure data used in an aggregate risk assessment?

To integrate drinking water data into an aggregate assessment, Monte Carlo or other probabilistic methods should be used. The selected method should sample from defined variability (of primary importance) and uncertainty (of secondary importance) distributions describing each exposure pathway, with incorporation of correlation between exposure pathways. It should consider both intersubject variability of exposure factors at a given age (including consideration of special defined populations with unique exposure activities) and exposure factor variability between specific age groups (e.g., neonates, young children, and adults). These methods for reflecting vari-

ability and uncertainty have been well developed in other guidance for risk assessments performed for regulatory decisions (EPA 1997b). There are, however, several obstacles to implementing these methods for the case of aggregate risk assessment considered here.

Most of the available data (e.g., NAWQA) are collected for other purposes and often are not from sources used for drinking water. Other data do not focus on tap concentrations (i.e., the concentration at the tap in a home) but, rather, on the original water supply (e.g., the concentration after treatment but before distribution). The resulting distributions of pesticide concentrations therefore do not incorporate the effects of distribution and mixing. In many cases, they do not even incorporate the effects of treatment. In addition, much of the information in the NAWQA data set were not obtained at the point of intake into the water treatment system but, rather, at points "upstream." They do not, therefore, account for changes in concentration that might take place between the sampling point and the point of intake into the treatment system (e.g. owing to sedimentation or chemical reactions). It will be necessary to account for these effects in the future, since the data on other exposure pathways do allow incorporation of analogous factors (e.g., food consumption can be based on market basket studies that include the effects of treatment and distribution systems). In some cases, it is likely that the difference between concentration in finished water and that at the tap will not be large, so the lack of existing tap data may not be a significant source of uncertainty if finished water data are available. Use of source water as a surrogate for tap water will, however, introduce significant uncertainties into aggregate assessments. It will be necessary to apply correction factors to account for treatment.

The existing data generally oversample high exposures in a region, since a water supply with a high value is likely to be sampled repeatedly to ensure compliance or to isolate the cause of a problem. This issue is not as prevalent in the databases available for the other pathways. This limits the utility of a pooled sample of water supply measurements to improve the characterization of the population of supplies (sampled and unsampled) within the region, since even the pooled sample will overrepresent the supplies yielding high exposures. As a first step, it is recommended that samples for a given supply be pooled first to estimate concentrations at that supply, and only then would the summary statistics (e.g., means from the different supplies) be examined to establish the variability across the larger (sampled plus nonsampled) population. This would prevent a supply with many samples from dominating the pooled data. The exception to this rule would be cases where it was determined that sampling was random and representative.

Correlation between exposures through different exposure routes has not been determined to date. At least for the present, it must be assumed that exposure pathways are uncorrelated. This assumption is likely to be less valid for small, special populations such as subsistence farmers. Some measurement of temporal correlation between exposure pathways will be needed, particularly for the short term (e.g., 1 day) exposures associated with acute effects. By at least accounting for correlation during use seasons, the effect on uncertainty can be minimized. In addition, exposures through the various routes tend to be through environmental media obtained from diverse geographic regions. For example, the air that individual people breathe, the food they consume, and the water they ingest often are from different locations. This tends to remove some of the correlation that would exist had these environmental media all been obtained from the same geographic location. The exception will be for special subpopulations such as subsistence farmers, who often make up the maximally exposed population that drives regulatory decisions.

In some cases, trends will be noted in the temporal data for a supply. In taking time-weighted averages, one should examine these trends because they may indicate increasing concentrations (owing to the recent introduction of a pesticide into the watershed) or decreasing concentrations (owing to washout of a pesticide no longer in use). It is important that these trends be examined under a trend test and that the trend, if present, be factored into projections into the future. If no trend is noted, the time-weighted average may be determined directly from the data. The issue of trends will be particularly important during times of product introduction or as a ground water plume passes the point of intake for a well. Trends will be less significant once product use is fully established. They also will be less significant when assessments are performed across pooled water supplies, since temporal increases in one supply may be offset by decreases in another (as crop patterns shift within a geographic region).

Development of the full, and unbiased, variability distributions for drinking water exposures is limited by sparse data. This limitation can be partially removed by considering a form of meta-analysis of the existing data sets, rather than relying solely on one data set. In performing such an analysis, one must determine whether a given data set represents a true random sample over some defined population or whether it is biased toward the more contaminated supplies (a characteristic of many of the data collected for regulatory purposes). If the latter is true, it is recommended that the data be used only to define the upper tail of the variability distribution rather than assuming it applies to the entire variability distribution.

4.4.2. Given existing data, how can probabilistic distributions for intersubject variability be generated for different assessment scales (a national distribution, a regional distribution, and a local-scale/highly vulnerable water supply) within the pesticide's area of use?

A consistent methodology must be applied in moving from the site-specific, to the regional, and to the national scales. These three scales yield the same methodological issues, with the regional being a composite of the site-specific data and the national being a composite of the regional data. There is no reason, therefore, why the characterization of variability of exposure through drinking water should differ between the scales. The sole caveat is that the site-specific data usually are generated for a highly vulnerable supply, rather than representing a random selection of supplies from within a region. Data generated for the assessment of highly vulnerable supplies therefore will not form an appropriate basis for estimating the variability of exposures across a region, although they will form a basis for the upper-bound estimate of exposures in the region (as might be useful in a screening assessment). The problem with using such data as a surrogate for the regional exposures is that they do not, by themselves, provide an understanding of where those supplies lie in the variability distribution for the region, so it is not possible to state quantitatively the degree of conservatism they introduce into the analysis without first examining a random sample of water supplies from the regional sample.

Although the highly vulnerable supply data do not form a proper basis for estimating regional (and hence national) exposures, they do provide information that should not be lost when moving to the regional or national levels of assessment. *It is recommended that the highly vulnerable site data be supplemented by a more random sample but not discarded, since the former will provide a more accurate picture of the upper tail of the variability distribution within a region than might otherwise be obtained from a completely random sample.* Appropriate statistical methods should then be used to pool of the targeted (highly vulnerable) and random samples.

It is not recommended, however, that the highly vulnerable site analysis be simply nested inside the regional analysis, or that the regional simply be nested inside the national. Although it is important to retain information gained at a lower geographic scale when creating the data set for a larger scale, there will be issues of representative samples that change when moving from one scale to another. *The most obvious example is in creating the data set for the highly vulnerable site analysis. These sites cannot be identified properly unless a reasonably large regional-scale analysis has been conducted and used as a partial basis for designating a site "most vulnerable."* Thus, in this case, the regional-scale data or model predictions play a role in selecting the highly vulnerable sites, rather than simply incorporating the latter.

For a given region, samples from a given supply should be pooled within each of the "use" seasons (i.e., one pool for season of application, one for periods of nonuse, etc.). Distributions of the supply-specific samples then should be analyzed for mean and variance. After that, the distributions (means and variances) for the different water supplies in a region should be assessed for comparability using the appropriate statistical test. If they are judged sufficiently similar, greater statistical power can be gained by pooling the samples from across supplies within a region. This pooled sample would be used to generate the variability distribution for the entire region and for each season. Significant errors in the analysis can be introduced, however, by failure to consider appropriate statistical procedures when combining data.

Variability in exposure factors (e.g., water ingestion rate per unit body mass) should be obtained from a common reference. The most relevant references are the *Exposure Factors Handbook* (EPA 1997a), which contains the necessary values divided by age group and activity pattern, and the Continuing Survey of Food Intakes by Individuals (CSFII) database, which has the advantage of also being used to provide data for the dietary portion of the aggregate analysis. For the site-specific cases (vulnerable supply), the intersubject variability caused by exposure factors will dominate consideration of variability. It is recommended that correlation between exposure factors for an individual be considered. For example, water ingestion rates and food consumption rates may be correlated, since these tend to scale with body mass. This correlation will be less significant if the exposure factors are placed on a "per body mass" basis.

4.4.3. What is the level of accuracy with which these variability distributions and/or point estimates can be generated using data?

Because of the fact that the existing data are not the result of a fully randomized and stratified sample, they will provide low to medium accuracy for the regional-scale assessment in regions where data have been collected. This accuracy is lower for the national-scale assessment because of smaller sample size in areas where pesticide water concentrations were not of direct regulatory concern (past sampling having been done primarily in response to regulatory issues rather than in response to the need for a

systematic, randomized, stratified sampling across the nation). This accuracy is higher for the site-specific case, since this case is designed to consider the most contaminated supplies, although even here there is a problem in ensuring that the site selected truly represents the highly vulnerable supply. This problem of identification is removed to some degree by focusing on a set of sites deemed to constitute the highly vulnerable sites, recognizing that in any one time period the single most highly vulnerable site may shift but should still be contained somewhere in the set.

It is particularly the case that the exposure associated with any identified percentile of the upper tail of the national distribution of pesticide concentrations in drinking water is likely to be estimated inaccurately unless it can be assumed either that all unsampled supplies are part of the same distribution as the sampled supplies (currently a poor assumption) or that they have concentrations significantly below the sampled sites. Even if this latter assumption is valid, however, the very large size of the population represented by unsampled supplies is likely to push the sampled population into the extremes of the upper tail for the exposure distribution over all (sampled plus unsampled) supplies, meaning that the national distribution might be characterized at these extremes (e.g., above the 99th percentile) but not in regions of the distribution likely to control regulatory decisions (e.g., in the 90th to 95th percentiles).

Improvements can be made in the accuracy of the upper tails by using models to fill gaps for supplies where data are not available. The accuracy of such models, if used uncalibrated, is low. By calibrating the models to existing data in a region, and then using the calibrated models to make predictions for unsampled supplies, one can improve estimates of the variability distribution across a region. Although the models are not judged to yield accurate absolute estimates of concentrations, there is moderate confidence that they can yield relative values of concentrations—hence the need for calibration. *An issue that will remain, however, is how the calibration is to be performed* (i.e., which parameters in the model are to be adjusted). Accuracy is likely to be higher (as will confidence) if the choice of parameters to adjust can be given scientific validity and/or if it can be shown that the predictions of exposure are relatively insensitive to the choice of adjustable parameter.

4.4.4. How well do existing data and model predictions represent actual exposures in the population at different assessment scales?

Existing data focus primarily either on supplies not used as drinking water, on the concentration in drinking water supplies before treatment, or on concentrations exiting the treatment facility. Significantly fewer data directly measure concentrations at the tap, and models currently do not focus on tap concentrations. Still, because the effect of the distribution system on pesticide concentration is likely to be small, it is a reasonable first approximation to assume that measurements and models that focus on treated water are representative of tap water concentrations for the sampled supplies.

The greatest problem lies with models used routinely in regulatory analyses for pesticides; more strictly, the problem lies in the selection of the scenarios to which the models are applied. Models such as PRZM/EXAMS (for surface water) and SCI-GROW [AU: Spell out these abbreviations—or at least cite them in the References?] (for ground water) provide highly conservative estimates of concentrations if highly conservative scenarios are applied and they are not calibrated; they also do not account for the effects of treatment, dilution, and distribution. At the moment, it is not possible to state where the results of such uncalibrated models lie within the variability and uncertainty distributions that characterize actual exposures. All that can be said is that the models produce results in the extreme of the upper tail of the variability distribution and of the uncertainty distribution when the scenarios typically applied in regulatory analyses are used, but no more quantitative assessment can be provided. This means that these models are not suitable for generating the variability and uncertainty PDFs discussed in Section 4.3 at either the national or the regional scale until such time as more realistic scenarios (and distributions of scenarios) are developed. It is possible, however, that such models might be suited to estimating exposures from a hypothetical highly vulnerable water supply under existing scenarios, although it should be confirmed that candidate supplies possessing the characteristics assumed by these models exist within a region before applying them. *If aggregate exposure is shown to be acceptable under these model results, it can be concluded with high confidence that aggregate exposure would also have been judged acceptable under more accurate estimates of the drinking water contribution.*

It is assumed here that risk management decisions should be based on at least medium accuracy for predictions of aggregate exposure to have a reasonable chance that policies will meet risk management goals. With this in mind, the following statements can be made:

For exposure assessments requiring low accuracy and low levels of confidence, existing data and models could be used to establish estimates of aggregate exposure for the upper tails of the distribution of pesticide concentrations in drinking water for site-specific and regional scales of assessment, but it will not be possible to accurately specify the percentile associated with these estimates. In

other words, the assessor will be able to state that the estimate of aggregate exposure is likely to be higher than that applicable to a large fraction of the population, but will not be able to specify whether it is higher than that for 95%, 99%, etc. It is likely that the estimate will be well above the 95th percentile.

If appropriate models are developed, properly validated and calibrated against monitoring data, and used to extrapolate from sampled water supplies in a region to unsampled supplies, it may be possible to establish an approximation of the full distribution of pesticide concentrations in drinking water in a region with low to medium accuracy.

To obtain high accuracy, a random, stratified monitoring program will be needed that includes a design component for oversampling in vulnerable sites, with information retained on population served, degree of treatment, type of source water, etc., for each supply.

4.4.5. How must data and model results be generated in the future if they are to represent actual exposure conditions in the population?

Optimally, there must be a regional (and through combining of results, a national) randomized and stratified sampling performed of treated water (where treatment exists; otherwise the source water is sampled). This sampling should include the ability to estimate 1-day averages (for acute effects) and annual averages (for chronic effects). Given the likely subsequent calculations that will need to be performed, and given risk management goals, it is recommended that sampling be stratified on at least most of the following characteristics:

- use season (needed to account for correlation in exposure pathways),
- use area,
- vulnerability,
- treatment process (needed to correct source water to finished water), which may be replaced by some surrogate that combines size of water supply and rural/urban distinctions, and
- special subpopulations of regulatory interest (e.g., tribes, minority groups, etc.).

Until such a sampling is performed, it is unlikely that estimates of the variability distribution for exposures through drinking water will be judged to be of more than low to medium accuracy. With such a sample, the accuracy should be raised to medium or high. If resources permit, the ideal situation would be to perform a study on tap water (rather than source water or finished water) using the same study design, although gaining the necessary access for samples might be difficult. The sole caveat to performing a study only on tap water is that it may preclude use of the data in identifying alternative methods for exposure reduction (such as watershed protection).

As for models, there is a great deal of improvement needed in both the models and the scenarios to which they are applied before their results will provide an accurate estimate of actual exposures. At present, the models used represent only maximal credible concentrations, not concentrations useful in exposure assessment and risk assessment. Future models must be based on more realistic modeling of watersheds and be linked to Geographic Information Systems (GIS) and associated data if they are to provide a basis for estimating relative exposures from different supplies. Their role is likely, in the short term, to remain one of extrapolating and interpolating between data rather than for generating a priori estimates of concentration at a supply. The use of models to perform such extrapolations will require that the utility of those models be confirmed.

4.4.6. Under which conditions is it appropriate to extrapolate exposure data temporally to support an aggregate assessment?

As described previously, it will be essential to test for trends in data before extrapolating in time. If the daily variability distribution for a supply has been generated for a sufficient period (long enough to include variations in climate, pesticide use, etc.), and if the distribution in daily values across those individual years is judged by the appropriate statistical test to be stable, and if rainfall during this period has been shown to be representative, the use of these data directly for future years will be justified *until such time as the risk assessor determines that use patterns have changed significantly*. If the distributions are changing in time, this trend should be characterized and either extrapolation is not performed (if the trend is significant and if the cause of the trend is completely unknown), the trend is ignored and the uncertainties are noted through broadened confidence intervals, or the trend is characterized quantitatively and the distribution is allowed to evolve in time in a manner consistent with that trend (e.g., adjusting the mean of the distribution over time). For at least the case of ground water contamination, the identification of trends should be based on an understanding of the travel time of the contaminant within the aquifer. If the condition on travel time has not been met (i.e., the time since contamination is smaller than the travel time from the point of contamination to the point of sampling), it should not be assumed that prior sampling is representative of future exposures.

4.4.7. Under which conditions is it appropriate to interpolate data spatially within a region and to extrapolate spatially from one region to another?

It is recommended that these interpolations and extrapolations be performed only under one of two conditions:

(1) A properly randomized and stratified sample from the water supplies in a region has been performed (existing data sets generally do not satisfy this assumption). In this case, the data set obtained can be used to represent the unsampled water supplies in the same region (i.e., the measured variability distribution itself is extrapolated unchanged to the unsampled sites that lie geographically between the sampled sites). It is important that the sample design reflect the fact that aggregate exposure assessments will need to perform weighting by population.

(2) A GIS-based model of ground water and surface water contamination (not one of the existing regulatory models, which are insufficient for this task) is used to predict relative values of the concentrations at the unsampled sites based on localized application rates and topographic features, and this model is calibrated to the data at the sampled sites for purposes of extrapolating to the unsampled sites.

4.4.8. What are the comparable features needed between regions to allow such extrapolation spatially?

If the extrapolation is done purely statistically (i.e., without modeling), the regions should have the same fractions of individuals drawing from surface and ground water supplies (or the difference must be corrected for by appropriate weighting factors); should have approximately the same fraction of land devoted to crops that involve routine use of the pesticide (although it may be possible to adjust for this fraction); should have approximately the same climatological conditions (although it may be possible to adjust for differences in rainfall); should have approximately the same application practices for the pesticide; and should have comparable treatment and distribution systems (or correction for subsequent treatment and distribution can be applied).

To reduce uncertainty, the extrapolation could be guided by the GIS-based advanced models for surface and ground water contamination discussed previously. This would require detailed information on pesticide use, topography, water supply locations, etc., for both the region of interest and the region from which estimates are to be extrapolated. The advantage of this approach is that it uses the models to predict how the variability distribution from the data-rich region is to be adjusted in extrapolating to data-poor regions, relying on the ability of the models to provide relative (but not absolute) estimates of exposure in a region. *In any case, data from one region should never be simply used in another region (or at another scale) without fully considering the differences between the regions.*

4.4.9. For national and regional assessments, can distributions of pesticide concentrations from different water supplies be combined? Under which conditions? Using which methodology?

It is not only useful but necessary to combine results from water supplies in developing the national and regional assessments (the exception is the pooling of highly vulnerable site data to produce the regional assessment, which is limited by concerns raised earlier). The reason is that many of the water supplies in such regions are not sampled, and the sample size is small for any given supply. Accurate characterizations of the variability distributions for a region (and hence the nation) will require pooling results from different supplies.

Whenever this is done, however, it must be assured that the water sampled is at the same level of treatment in all of the pooled samples (i.e., all are primary water, all are finished water, or all are tap water) and that the temporal variability distribution of concentrations at a given water supply is consistent with the distribution from each other water supply placed into the pooled sample. This consistency should be tested quantitatively using an appropriate statistical test to ensure that the data are being drawn from approximately the same underlying population. Where this is not the case, pooling should not be performed if individual risk is of interest. If only population risk is of interest, pooling may be performed even when the temporal distributions are dissimilar.

In addition, it must be ensured that data from the regions to be pooled were collected under the same sampling regime and summarized in the same manner before being pooled. Each water supply/region should contribute to the national sample only to the extent that supply serves a fraction of the total population (e.g., a water supply serving 200 people should count twice that of a supply serving 100 people in any Monte Carlo analysis). This is accomplished by summarizing the temporal data from a given supply as a temporal variability distribution and weighting that variability distribution by population in the Monte Carlo analysis, rather than simply pooling data from all water supplies into a composite variability distribution. Note that the population weighting is performed after data sets have been collected, rather than being used in the collection of the data themselves. *However pooling is treated, it is essential that the pooling be done under defined statistical procedures, because significant er-*

rors can be introduced through improper pooling.

4.4.10. What is the appropriate temporal scale over which drinking water exposures should be summarized so that these exposures can be united with exposures by other routes?

Three temporal scales are needed:

- Daily time-weighted averages are needed for acute effects.
- Annual time-weighted averages are needed for chronic effects.
- Lifetime time-weighted averages are needed for cancer endpoints.

Given variations during seasons, it will be necessary to specify at least the 1-day averages for the use season and the nonuse season. It is unlikely that many of the other exposure pathways will have similar temporal information, so it is likely that more detailed water data will be combined with less detailed temporal data on the other pathways. Still, the accuracy gained by considering seasons for the water data justifies the additional work needed to keep season data separate. At the same time, the decision to collect additional data or develop more detailed models for estimation of daily exposures through water should be tempered by the limitations of information on other exposure pathways.

4.4.11. When is it appropriate to use time-weighted averaging for a single water supply?

The concentration of a pesticide in water ingested by an individual varies in time; i.e., there is temporal variability. In such a case, what is the appropriate water concentration to use in an aggregate assessment? If only the average concentration is used, information will be lost on peak concentrations that might be responsible for most of the risk, even though they occur only a fraction of the time. If only the single highest concentration is used, this might lead to an overestimate of the risk in cases where that peak concentration is temporary and does not exist long enough to cause any effect. The appropriate summary of the concentration is to determine the average concentration over some period of time, such as a day, year, or lifetime; the specific time period chosen should match the minimal period of time a sufficiently high exposure must be maintained to produce an effect.

Temporal variability, therefore, raises the issue of selecting an appropriate temporally averaged exposure of a population using that water supply. For chronic and cancer endpoints, at least under an assumption of linearity between exposure and response (the traditional approach in regulatory risk assessment), the time-weighted average concentrations in the various exposure media (including drinking water) should be used. For each water supply, the temporal variability would be averaged over the appropriate time interval, with random selection of the starting point for these time intervals across a time series of results.

The discussion above assumes that the concentrations in a water supply vary randomly in time. In some cases (such as ground water supplies with plumes of a pesticide migrating through the aquifer), there will be a distinct trend in time, for example, generally increasing and then decreasing slowly as a plume passes the monitoring point. If this is the case, it will not be correct to select randomly from the pool of all data, but rather one should select only from those data taken during the period of time judged to be representative of exposures during the period of interest. The concept of temporal variability must also be taken into account when looking at the source of the water data. For example, if water is drawn directly from a small river, treated, and then passed along with minimal residence time in storage tanks to consumers, temporal variability in the tap water may closely reflect that in the source water. If, however, data are used that came from samples taken from a river that enters a reservoir used for the storage of drinking water, the temporal variability of the samples may be significantly different from that in the tap water, especially in terms of short-term concentrations.

4.4.12. What are the most significant assumptions, limitations, and uncertainties associated with data, model results, and aggregate exposure estimates?

By far the most important limitation is the lack of a fully randomized and stratified sampling of water supplies organized by region, season, and treatment/distribution (or surrogate). The existing databases provide only limited sampling, and their use is judged to provide only low accuracy in characterizing the variability distribution. There is moderate accuracy in characterizing some upper percentile of the distribution, but it cannot be stated accurately where specific percentiles fall in that distribution (e.g., where is the 95th percentile?). All that can be said with moderate accuracy is that the *X*th percentile of the measurements is likely to represent greater than the *X*th percentile of the true variability distribution owing to the likely oversampling of the more contaminated supplies (or at least those supplies judged likely to be contaminated). Such data should be used only for screening purposes (i.e., as an indicator that additional monitoring may be

necessary) of for selected site-specific analyses that have been characterized as to uncertainties and representativeness of the data.

Similar problems underlie the estimates of exposure for other pathways; taken together, this will compound the conservatism associated with the aggregate exposure assessment. The lack of data on correlation between exposure pathways, both geographically and temporally, is problematic but probably not of the order of concern introduced by the lack of a properly randomized sample.

With respect to the models, the major limitation is that the models and/or scenarios used currently are conservative, and the degree of that conservatism has not been estimated by comparing predictions against specific percentiles in the tails of the variability distributions. Such a comparison is recommended for the near future. In addition, the conservatism should be removed by adopting or developing more advanced, GIS-based models. The limitations in these models are data availability (particularly for the site-specific predictions) and the lack of validation/verification of the models. Such models may form the basis for estimating relative concentrations in water supplies; at present they are useful only when they are calibrated against data from the sampled supplies and the calibrated model is then used to make predictions for the unsampled supplies.

4.4.13. How should water treatment be accounted for in producing estimates of pesticide exposure through drinking water?

Some of the existing data involve measurements of concentrations in finished water, so additional consideration of treatment is not required; most do not. The problem lies primarily with the remaining data that focus on the source water or even simply potential source water, and with the use of models (since even the advanced models fail to incorporate treatment). To account for the effects of water treatment, it will be necessary to develop a national database on the ratio of treated to source water concentrations before the data on source water concentrations and the model results can be used to characterize drinking water exposures accurately. The relevant stakeholders (the EPA, American Water Works Association [AWWA], and pesticide registrants) would need to work together to develop a national database of the necessary conversion factors. *This database should be stratified by indicators of treatment/distribution such as size of population served, rural/urban and/or surface/ground water. The same stratification should be retained in any monitoring program for water concentrations.*

4.4.14. How should water distribution be accounted for in producing estimates of pesticide exposure through drinking water?

Two primary issues are identified: the water distribution system may change the concentration as a pesticide flows through the system because of chemical reactions and/or plateout, and the distribution system may include mixing of water supplies. With respect to the first issue, it is judged that the uncertainty in risk estimates for drinking water will not be increased greatly, since tap and finished water concentrations tend to be similar for most pesticides (Steven Via, AWWA, personal communication, 1998). With respect to the second issue, this will present a significant problem only in supplies where mixing is both common and extensive (i.e., two or more water supplies represent a significant fraction of the mixture). It is recommended that geographic regions where mixing is prevalent be identified and a representative mixture be assigned to each such region. At the moment, the necessary data do not exist, and it must be assumed that finished and tap concentrations are approximately equal.

References

Cohen J, Lampson M, Bowers T. 1996. The use of two-stage Monte Carlo simulation techniques to characterize variability and uncertainty in risk analysis. *Human and Ecological Risk Assessment* 2:939–971

Crawford-Brown D. 1997. *Theoretical and mathematical foundations of human health risk analysis.* Dordrecht, Netherlands: Kluwer Academic Publishers

Finley B, Proctor D, Scott P, et al. 1994. Recommended distributions for exposure factors frequently used in health risk assessment. *Risk Analysis* 14:533–553

International Life Sciences Institute. 1998. Aggregate exposure assessment. Washington, DC: ILSI Press

Morgan M, Henrion M. 1990. Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis. Cambridge: Cambridge University Press

US Environmental Protection Agency. 1990. National Pesticide Survey: summary results of EPA's National Survey of Pesticides in Drinking Water Wells. EPA570990NPS5. Washington, DC: EPA

US Environmental Protection Agency. 1996. Proposed guidelines for carcinogen risk assessment. *Federal Register* 61:17960

US Environmental Protection Agency. 1997a. Exposure factors handbook. EPA/600/P-95/002Fa. Washington, DC: EPA

US Environmental Protection Agency. 1997b. Guiding principles for Monte Carlo analysis. EPA/630/R-97/001. Washington, DC: EPA